



Cognitive Science 35 (2011) 399–443

Copyright © 2010 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/j.1551-6709.2010.01145.x

What a Rational Parser Would Do

John T. Hale

Department of Linguistics, Cornell University

Received 4 June 2009; received in revised form 30 July 2010; accepted 31 July 2010

Abstract

This article examines cognitive process models of human sentence comprehension based on the idea of informed search. These models are *rational* in the sense that they strive to find a good syntactic analysis quickly. Informed search derives a new account of garden pathing that handles traditional counterexamples. It supports a symbolic explanation for local coherence as well as an algorithmic account of entropy reduction. The models are expressed in a broad framework for theories of human sentence comprehension.

Keywords: Comprehension; Entropy reduction; Heuristic; Rationality; Sentence processing; Syntax

1. Introduction

As in other areas of cognitive science, the study of human sentence comprehension naturally proceeds at multiple levels. Marr (1982) draws a distinction between scientific questions posed at the *computational* level and questions at the *algorithmic* level. At his highest level, the computational level, one asks: What task does a sentence understander carry out? What sort of mental representations are constructed, as the outputs of sentence comprehension? What features of the comprehended language are brought to bear, as inputs to the comprehension task? Is this input-output mapping “easy” or “hard” when considered as a formal problem on its own? Marr’s distinction separates these concerns from others that pertain to the algorithmic level. At this lower level, one asks: What procedure accomplishes the calculation of the function defined at the computational level? What is the time course of this calculation, and what is its resource utilization in terms of abstract resources like storage? This separation into independent levels of analysis constitutes one of the most distinctive methodological features of cognitive science.

Correspondence should be sent to John T. Hale, Department of Linguistics, Cornell University, Room 217, Morrill Hall, Ithaca, NY 14853-4701. E-mail: jthale@cornell.edu

Marr's levels help classify the kinds of progress that have been made over the years in the field of sentence comprehension. From the empirical side, a growing number of information sources have been demonstrated to affect human performance. These factors include but are not limited to lexical bias (Ford, Bresnan, & Kaplan, 1982; Mitchell, 1987), referential status (Crain & Steedman, 1985), prosody (Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991), argument structure (Boland, 1993; Stowe, 1989), event prototypicality (McRae, Ferretti, & Amyote, 1997), and real-world knowledge (Chambers, Magnuson, & Tanenhaus, 2004). Work of this sort contributes an increasingly detailed specification of the inputs to comprehension. By characterizing the information that people evidently use, we converge on a more precise computational-level statement of the problem itself.

On the theoretical side, too, there has been progress at the computational level.¹ Information-theoretical complexity metrics involving entropy (Hale, 2003, 2006; Frank, 2010; Roark, Bachrach, Cardenas, & Pallier, 2009; Yun, Whitman, & Hale, 2010) and surprisal (Hale, 2001; Boston, Hale, Patil, Kliegl, & Vasishth, 2008; Boston, Hale, Vasishth, & Kliegl, 2011; Demberg & Keller, 2008; Levy, 2008) have been used to recharacterize the difference between “easy” and “hard” comprehension situations. Bayesian probability models have recast sentence comprehension as the revision of beliefs about a sequence of words (Narayanan & Jurafsky, 1998, 2002). And the work of Chater, Crocker, and Pickering (1998) and Crocker (2005) has reconnected this sort of “disinterested” information-processing with the concept of *utility*, which incorporates a quantitative notion of relevance to an agent's goals. This line of research explicitly invokes Anderson's (1990) *rational analysis* method in explaining comprehension. But like the other theories mentioned above, the analysis is pitched at the computational level, and thus, remains silent on questions of mechanism. In order to move ahead, the field's progress at the computational level should be connected to proposals at the algorithmic level. The next generation of sentence comprehension theories should be able to incorporate the diverse information sources that humans evidently use while at the same time providing insight into the cognitive process that is, fundamentally, the object of study.

1.1. The proposal and its significance

This article introduces a broad, algorithmic-level framework for theories of human sentence comprehension. The framework synthesizes two recurring themes in the psycholinguistics literature. One is that comprehension can be understood as a process of applying grammar rules. Another is that the choice between applicable rules is governed by broader rationality principles. Each of these themes has been extensively explored on its own. For instance, the relationship between grammar rules and perceptual processes is a central topic in work on Augmented Transition Networks (Kaplan, 1972; Thorne, 1968) as well as the Marcus parser (Marcus, 1980). The idea of ambiguity resolution based on experience, a kind of rationality principle, has been taken up in Constraint-Based Lexicalism (MacDonald et al., 1994) and as part of the Tuning Hypothesis (Mitchell, Cuetos, Corley, & Brysbaert, 1995). But perhaps because each idea offers so much, on its own, interactions between the

two have only rarely been studied with implemented models. This article's synthesis brings these interactions into sharp relief. It reconceives parser heuristics, such as Late Closure, as adaptive responses to explicitly defined conditions. It treats "serial" and "parallel" processing as consequences of more fundamental pressures. It allows a grammatical view of intermediate parser states, while at the same time relaxing the requirement that such content always remains self-consistent. The overall message is that comprehension can indeed be treated at the algorithmic level. Doing so sheds light on a set of questions whose answers are sometimes taken on faith by cognitive modelers. The proposal's significance thus reflects its high degree of engagement with fundamental issues in the theory of human sentence processing.

1.2. Organization of the article

The article combines the two themes mentioned above in a broad framework for comprehension. Section 2 introduces this framework. This section formalizes the idea of *applying* a grammar rule in terms of generalized left-corner (GLC) parsing (Demers, 1977). It presents informed search as a control regime for rule-applying systems such as GLC parsers. Under informed search, a comprehender chooses the action that minimizes cost. This section brings in A* search (Hart, Nilsson, & Raphael, 1968) as a way to optimize these costs.

Section 3 goes on to consider perhaps the simplest proposal within this framework. This simple model chooses its operations based on past experience in a literal way. It uses a parsing strategy based largely on traditional wisdom and naively adopts the syntactic annotations given in the Penn Treebank (Kučera & Francis, 1967; Marcus, Santorini, & Marcinkiewicz, 1993). Rather than attempting to fit a detailed model to complex data immediately, we consider the smallest possible system in which interactions between the two thematic ideas might occur.

Section 4 applies this simple model in accounting for three comprehension phenomena: Garden-path sentences (Frazier & Clifton, 1996; Frazier & Fodor, 1978), counterexamples to Garden Path Theory (Gibson, 1991), and local coherence (Konieczny & Müller, 2006, 2007; Tabor, Galantucci, & Richardson, 2004). In the first two examples, the model helps to explain why heuristic attachment strategies perform so well in some cases and yet must, in principle, derive the wrong prediction in others. The third example accounts for human performance in terms of intermediate states whose existence is implicated by rational analysis. This kind of explanation is uniquely available at the algorithmic level.

Results of this sort motivate an investigation of the simple model itself in Section 5. This investigation characterizes the operative principle in a way that does not depend on a Treebank. This principle can then be used in the definition of a "refined" model based on weighted grammar rules. These rules may be chosen to reflect any substantive linguistic theory. Section 5 exercises this flexibility by using a non-Treebank grammar to derive the asymmetry between mild and severe garden paths reported in Sturt, Pickering, and Crocker (1999). Sturt et al.'s asymmetry is a case in which two theories at different Marr-levels both derive the observed phenomenon. Closer study of this case points to a general way in which

the proposed algorithmic framework serves as a lower level realizer of a computational level proposal called the Entropy Reduction Hypothesis (ERH) (Hale, 2003, 2006). Section 6 contrasts the framework with a variety of other approaches besides the ERH, and Section 7 concludes.

2. Broad framework

Certain basic properties of the human sentence-processing mechanism are well established. To consider just three, it is widely agreed that human sentence processing is *incremental* in the sense that meaningful interpretation does not significantly lag behind the perception of individual words (Marslen-Wilson, 1973; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Almost as uncontroversial is the suggestion that language understanding is *predictive*. To say that a sentence understander is predictive is to suppose that it forms explicit representations of words and phrases that have not yet been heard. Within linguistics, it is also generally assumed that sentence understanding involves the recovery of a structural description for the given sentence. Such a structural description characterizes a particular sentence in terms of general properties of the language. This *competence* hypothesis (Chomsky, 1965, p. 9) is arguably the simplest way to unify linguistics and psycholinguistics (Stabler, 1991, p. 199), and it can be applied in stronger and weaker forms (Berwick & Weinberg, 1984; Bresnan, 1982; Steedman, 1989).

Even after ruling out proposals that are nonincremental, unpredictable, or irreconcilable with linguistic theory, there still remain an infinite number of programs that all calculate the same relationships between grammars, sentences, and their structural descriptions. This is a specific case of the general problem of *identifiability* in cognitive science. In flexible cognitive domains like language comprehension, it is challenging to locate evidence that substantially cuts down the space of possible mechanisms. Pylyshyn (1999) argues that this identifiability problem cannot be overcome empirically, even with intermediate-state evidence like verbal protocols or neurological signals.

Anderson (1990) offers rational analysis, summarized in Table 1, as a way to head off the identifiability problem. Although cognitive processes like language comprehension cannot be identified by their behavioral or neural effects, broader consideration of the task the

Table 1
Rational analysis (Anderson, 1990, p. 29)

1.	Precisely specify what are the goals of the cognitive system
2.	Develop a formal model of the environment to which the system is adapted
3.	Make the minimal assumptions about computational limitations...To the extent that these assumptions are minimal, the analysis is powerful
4.	Derive the optimal behavioral function given items 1 through 3
5.	Examine the empirical literature to see if the predictions of the behavioral function are confirmed
6.	If the predictions are off, iterate

system is solving provides an additional meta-theoretical constraint, which leads the way to a specific mechanism.

Although the identifiability problem still exists in principle, a successful rational analysis neutralizes it by yielding a result that is simple and makes sense with respect to the problem the cognitive system is solving. The broad framework introduced in this article reflects Anderson's rational analysis, and subsequent subsections take up each of the steps prescribed in Table 1.

2.1. Goals of sentence comprehension

For communication to happen, a sentence understander must work out what a speaker means. Achieving this goal is hampered by properties of a language like obliqueness, ambiguity, and vagueness. At the same time, verbal communication occurs under time pressure. Those of modest ability in a foreign language who have tried to participate in full-speed discussions with native speakers will recall the unpleasant feeling of being unable to keep up. These two considerations motivate the goal specification in Assumption 1.

Assumption 1.

The human sentence comprehension system's goals include rapidly arriving at the syntactic structure intended by the speaker.

The broad framework boils the goals of sentence comprehension down to two competing considerations: be accurate and be quick. In the context of the competence hypothesis, one can simplify the problem by supposing that these pressures affect the search for syntactic structure as opposed to some other level of linguistic analysis. This reflects the standard assumption (e.g., Bach, 1976) that semantic interpretation depends on syntactic structure, rather than any kind of denial of the relevance of nonsyntactic factors.

In the context of Anderson's (1990) methodology, the goal specification in Assumption 1 sets up a rationality principle: It would be rational to take actions that quickly lead to the interpretation the speaker intends. Such a principle is immediate because it is rational to take actions that lead to goals (Newell, 1980, p. 151). However, this formulation does not characterize how goals are actually achieved. In particular, this formulation does not assert that comprehenders are "ideal observers" who always succeed in using all knowledge to maximum efficiency.

2.2. Environment to which comprehension is adapted

Sentence comprehension always occurs in a speech community, a group of individuals who know the same language. Describing the knowledge that is shared by members of a speech community is one of the central goals of generative grammar. The results of this research—generative grammars—are themselves formal models of the environment to which comprehenders must be adapted. This view, itself a reflection of the competence hypothesis, is stated below in Assumption 2.

Assumption 2.

Human sentence comprehension is adapted to categorical and gradient information specified in the grammars of particular languages.

The broad framework itself does not include a claim about which grammar is best or where grammars come from. Rather, it asserts an identity relationship between the grammar that linguists write to describe a language and the expectations to which comprehenders are attuned.

2.3. Computational limitations

Assuming some generative grammar, a comprehension mechanism must additionally define a way of applying the rules of that grammar, as well as a way of resolving conflict when more than one rule is applicable (Steedman, 2000, p. 226). These are two of the most consequential design decisions in a computational model of human parsing. As regards the first computational limitation, the broad framework conceptualizes parsing as search in a *problem space* in the sense of Newell and Simon (1972). Transitions between states within this space are accomplished by the application of a single grammar rule. This respects a strong form of the competence hypothesis. As regards the second computational limitation, the broad framework supposes that parsing is guided by analysis quality and nearness to the completion of the comprehension task. This claim assigns a particular weak method (in the sense of Laird & Newell, 1983) to one potentially modular cognitive faculty (in the sense of Fodor, 1983). The next two subsections develop these proposals, answering the questions “what does it mean to apply a rule?” and “what would it mean to be closer or farther away from finishing?” respectively.

2.3.1. Parsing strategies

The broad framework adopts GLC parsing as a theory of transitions between problem states. While GLC has been extensively studied in formal language theory (for a review see Grune & Jacobs, 2008, §10.1.1) its obscurity in the sentence-processing community motivates a brief review here. The key idea is that particular rules of a grammar can be used to update an analysis in ways that could be more predictive or less predictive, and that this distinction differentiates various different strategies. For instance, top-down parsing “expands” a syntactic category such as VP, replacing it with two new categories V and NP as shown in Fig. 1. In Fig. 1, the symbols VP, V, and NP have brackets around them. This notation indicates they are *predictions* about material that has yet to be processed. On the top-down strategy, to “use,” “apply,” or equivalently “announce” a rule means to exchange a prediction for one category, for example, the left-hand side symbol VP, for a sequence of other categories such as V and NP. This predictive aspect of top-down parsing makes it attractive to cognitive modelers as there is evidence for these sorts of anticipatory representations (for a review see Kamide, 2008).

The alternative to top-down parsing is bottom-up parsing. A bottom-up parser exchanges evidence about a sequence of daughter categories for a conclusion about a parent category.

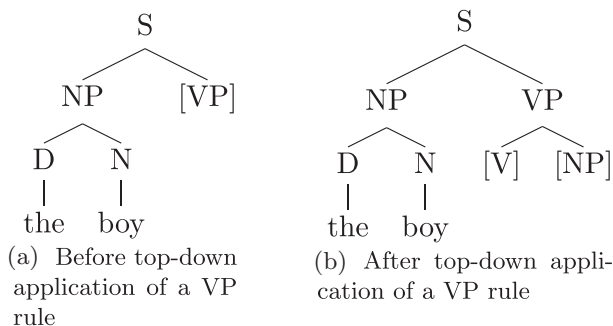


Fig. 1. The top-down parsing strategy applying the transitive verb phrase rule $VP \rightarrow V NP$.

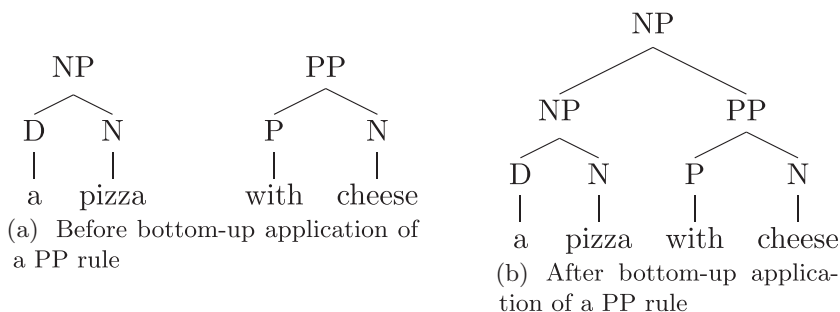


Fig. 2. The bottom-up parsing strategy applying the adjunction rule $NP \rightarrow NP PP$.

This exchange is shown in Fig. 2. In bottom-up parsing, there is no prediction. The lack of any bracketed categories in Fig. 2 reflects this absence of prediction. For instance, the larger NP is only built if in fact an entire PP is found.

Surveying these two strategies, Johnson-Laird (1983) concluded that the human comprehension mechanism must not use either of them. He offered a third way called the left-corner strategy as a psychologically plausible one. This strategy had previously been studied but not applied in models of human cognition by Rosenkrantz and Lewis (1970). The left-corner strategy is halfway between top-down and bottom-up parsing. If the first category on the right-hand side of a grammar rule can be found bottom-up, then the rule can be "applied." If there are other daughters specified on the right-hand side of the rule, they become predictions just as in top-down parsing. Fig. 3 exemplifies this strategy. In this example, evidence for the left-corner category Adj is exchanged for evidence about a bigger constituent, AdjP, whose subconstituent PP remains to be found.

GLC generalizes left-corner parsing by adding flexibility about the number of daughters that need to be found before left-corner projection, of the type illustrated in Fig. 3, can happen. Fig. 4 illustrates how each rule is associated with a number i . Daughters 1 through i are called the *trigger* of a rule; they must be found in a bottom-up way before the rule can be announced. After it is announced, the remaining nontrigger elements of the rule become expectations. Bottom-up parsing is then a GLC strategy where $i = n$, in other words where

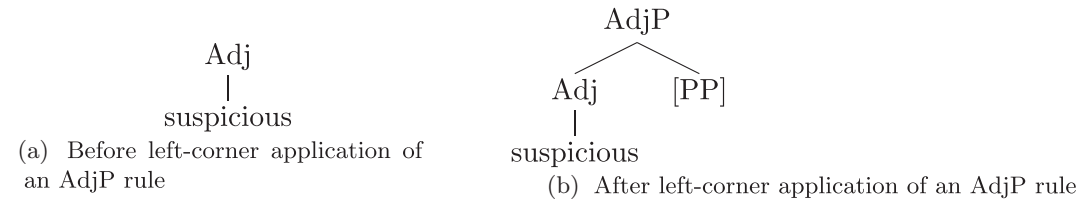


Fig. 3. The left-corner parsing strategy applying the rule $\text{AdjP} \rightarrow \text{Adj PP}$.

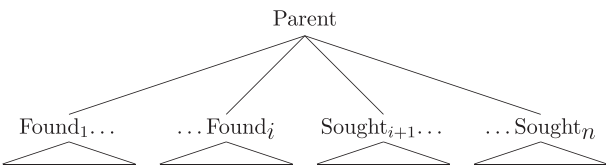


Fig. 4. Generalized left-corner parsing.

the trigger encompasses all of the daughters specified on the right-hand side of a phrase structure rule. Left-corner parsing is a GLC strategy where $i = 1$. And top-down parsing is a GLC strategy where $i = 0$. The GLC theory additionally offers the flexibility to assert that different grammar rules have different trigger points. Such “mixed” parsing strategies allow a cognitive model to treat arguments and adjuncts differently, as humans apparently do (for a review see Tutunjian & Boland, 2008). The choice of strategy itself constitutes a substantive parameter in the definition of sentence comprehension models.

Table 2 schematizes the broad framework’s proposal about possible transitions between problem states in parsing. Each transition fits into one of four possible categories. These categories represent two binary contrasts: A transition can either fulfill an expectation or not, and it either deals with a word or with a phrase. Within these categories, operations are parameterized to reflect the particular rule being applied. A particular GLC strategy identifies, for each rule, which daughters are part of the *Trigger* sequence that must be found before a rule may be “applied,” “announced,” or simply “used.” All of these terms mean the same thing within the framework.

Table 2
Four schemas define possible state-transitions

word vs phrase	
satisfies an expectation?	shift a word W
	project a rule $LHS \rightarrow Trigger$ $Rest$
	project and match the sought parent LHS
	scan the sought word W using the rule $LHS \rightarrow Trigger$ $Rest$

2.3.2. Informed search

When more than one transition is applicable, conflict arises. Should a comprehender project the transitive verb phrase rule, the intransitive one, or simply go on to the next word? Conceptualizing this problem as one of *search* brings to bear a wealth of theory and techniques from artificial intelligence (AI). From the classic AI perspective, a problem looks like a tree (Newell, 1968). Each node of this tree corresponds to a problem state. These nodes include all of the information needed to proceed with some line of reasoning. In GLC parsing, states must contain a stack of sought and found categories because this information conditions their applicability. The transitions between nodes in this search tree correspond to choices drawn from Table 2. Although the “project” operation applies a phrase structure rule, “shift” and “scan” deal with individual words either by pushing them on to the stack or popping them off, respectively.

The AI tradition distinguishes between search procedures that do not require any knowledge of the domain (“uninformed” search procedures) and procedures that do take advantage of problem-specific knowledge (“informed” search procedures). The broad framework supposes that human sentence comprehension involves the latter kind. That is to say the decision about which state to visit next is made on the basis of some notion of how good or bad the destination problem state is, rather than some fixed policy about the priority of specific transitions. If a searcher can work out the heuristic value $\hat{f}(n)$ of a node n , this quantitative notion of goodness can lead to better solutions sooner. To explore a problem space in this way, in order of heuristic value, is known as “best-first” search. One can take much of the literature on ambiguity resolution as being directed toward finding appropriate ingredients of the function $\hat{f}$ as suggested in Fig. 5. Increasingly complex models defined within the broad framework could carry out informed search using any number of information sources as long as they can be bundled together into a single heuristic value.

For better or for worse, the ordering of words in sentences makes it impossible to define accurate heuristic values for incremental parser states. This is because initial fragments of sentences can often be continued in a wide variety of ways. A particular syntactic analysis of the first few words of a sentence may be very good if it turns out to be followed by a certain continuation, but very bad if followed by another. The impossibility of a truly accurate valuation of incremental parser states follows from the fact that part of the problem involves guessing what the speaker intends to say in the future. In light of this, the framework applies a refinement of best-first search called A* search (Hart et al., 1968). In A*, the heuristic quality measure is split up into two kinds of costs as shown below in Eq. 1.

$$\hat{f}(n) = g(n) + \hat{h}(n) \quad (1)$$

One of these addends, $g(n)$, reflects the cost of a path from the initial state all the way to state n . The other, $\hat{h}(n)$, estimates the cost required to get from n to a goal state. The A* search procedure trades off these two kinds of costs in an effort to achieve a good solution overall without wasting too much time exploring unpromising subspaces.

It can be shown mathematically that if $\hat{h}$ satisfies certain conditions, A* search finds an optimal solution as fast as any search algorithm can. One of these conditions, “admissi-

$$\hat{f}(n) = Q(\text{definiteness}(n), \text{familiarity}(n), \text{prosody}(n), \text{felicity}(n), \\ \text{weight}(n), \text{memory burden}(n), \text{event typicality}(n), \dots)$$

Fig. 5. The broad framework combines analysis quality factors into a single cost, $\hat{f}(n)$. The precise inventory of factors and their mode of combination Q remain open questions in sentence processing.

bility,” amounts to the requirement that $\hat{h}$ be as conservative as possible. The broad framework does not impose this stringent requirement. From a cognitive perspective, it is more interesting to consider average-case expectations. In the following, we continue to refer to search guided by Eq. 1 as A* search even when $\hat{h}$ does not satisfy admissibility because of the way $\hat{f}$ is split up into the progress-so-far function g and the still-remaining estimate $\hat{h}$.

A* search thus provides a natural formalization of Assumption 1. The $\hat{h}$ term realizes Assumption 1's imperative to proceed rapidly, wasting no time on analyses that seem to be going nowhere. This imperative combines additively with the requirement to find a *good* analysis of what the speaker said. The g term formalizes this pressure.

2.4. Breadth of the framework

The proposed broad framework uses GLC and A*, respectively, to formalize the view: (a) that sentence comprehension uses grammar rules; and (b) that sentence comprehension is guided by broader rationality principles. This formalization yields a class of models that respect *incrementality*, *predictiveness*, and the *competence* hypothesis. Beyond that, the broad framework exerts relatively little constraint. For instance, no limits are placed on the notion of path cost as applied in A* search. This anticipates further discoveries about the factors that do and do not influence ambiguity resolution. Similarly, the framework does not make a claim about the substantive content of grammar rules. Consistent with Assumption 2, this anticipates the formulation of increasingly restrictive generative grammars for particular languages, including lexicalized grammars (Eisner, 2002; Schabes, 1990).

3. Optimal behavioral function: A simple model

Adopting the assumptions of the broad framework, this section lays out a particular comprehension model. To facilitate deeper analysis, this model abstracts away from several well-known aspects of human performance. It takes an intentionally simplified position on issues like adjunct processing, generalization from past experience, and lexicalization. It deploys just one information source, syntactic knowledge. This simplification allows the analysis to focus on two key parameters: the parsing strategy and the heuristic values $\hat{f}(n)$.

3.1. A mixed strategy

Inspired by work like Clifton, Speer, and Abney (1991) and Hemforth, Konieczny, and Strube (1993), the simple model adopts a mixed parsing strategy. This strategy applies all

rules in a left-corner manner, except for adjuncts. To put this same assertion in GLC terms, the trigger point is set to $i = 1$ for all rules except for those of the form $XP \rightarrow XP YP$ where XP has the same category in both the parent and the first child. Rules of this form are typically used in English grammars for optional postmodifiers. This simple mixed strategy calls for these to be parsed bottom-up. That is, the announce point i is set to n , the daughter-number of the final element of the right-hand side of the rule. This strategy has the salutary effect of relieving a parser from having to guess whether or not every noun phrase will be postmodified. Fig. 6 shows how this strategy applies toward the end of the globally correct analysis of Bever's (1970) famous garden-path sentence. In subfigure (c) of Fig. 6² the reduced relative clause *raced past the barn* has been analyzed as a verb phrase. This way of characterizing the phrase structure of relative clauses reflects the Treebank annotation guidelines (Marcus et al., 1993). From the perspective of the parsing strategy, the important point is that the globally correct reading calls for an adjunction rule. The reduced relative clause is adjoined to the noun phrase it modifies. The mixed parsing strategy does not predict these modifiers. Rather, once they are located bottom-up their mother node—in this case NP—is projected. This bottom-up handling of an adjunct is shown in subfigure (d), which is analogous to Fig. 2(b). Having built a complex subject NP, the analysis can proceed by projecting $S \rightarrow NP VP$ in a left-corner manner (subfigure (e)) creating an expectation for a VP which is eventually fulfilled by *fell* (subfigures (g) and (h)).³

This strategy can handle disconnected pieces of phrase structure. These are separated by semicolons in Fig. 6 and each successive position may be thought of as a kind of memory cell. Abney and Johnson (1991) and Resnik (1992) present the argument that this strategy avoids unrealistic storage requirements when applied to English sentence structures. Work by Schuler, Miller, AbdelRahman, and Schwartz (2010) suggests that on a predominantly left-corner strategy like this one, as few as four memory cells are sufficient to analyze 99% of the *Switchboard* and *Wall Street Journal* corpora. On the other hand, as a reviewer points out, this strategy does not allow adjuncts to be added to a category that has already fulfilled an expectation. This classic problem has motivated a variety of techniques for underspecifying the phrase structural content of intermediate parser states (Demberg & Keller, 2009; Shieber & Johnson, 1993; Sturt & Crocker, 1996; Sturt, Costa, Lombardo, & Frascioni, 2003; Weinberg, 1993). One or another of these techniques could be adopted in an elaborated version of the model.

3.2. An empirical search heuristic

In the simple model, the heuristic values that guide informed search are *costs* expressed as numbers of GLC steps. This idea, that language users favor shorter analyses, has a distinguished history. It has been advocated by theorists such as Noam Chomsky (1995, p. 145), Joan Bresnan (2001, p. 91), and Lyn Frazier (1979, p. 36). Paul Grice's (1975) Maxim of Manner could be rephrased to say that a cooperative speaker's intent is reflected in the shorter analysis. While it remains an open question exactly which notion of heuristic quality guides the human sentence processing mechanism, the simple model reifies the assumption that brevity will turn out to be a key component. In the context of A* search, this idea can

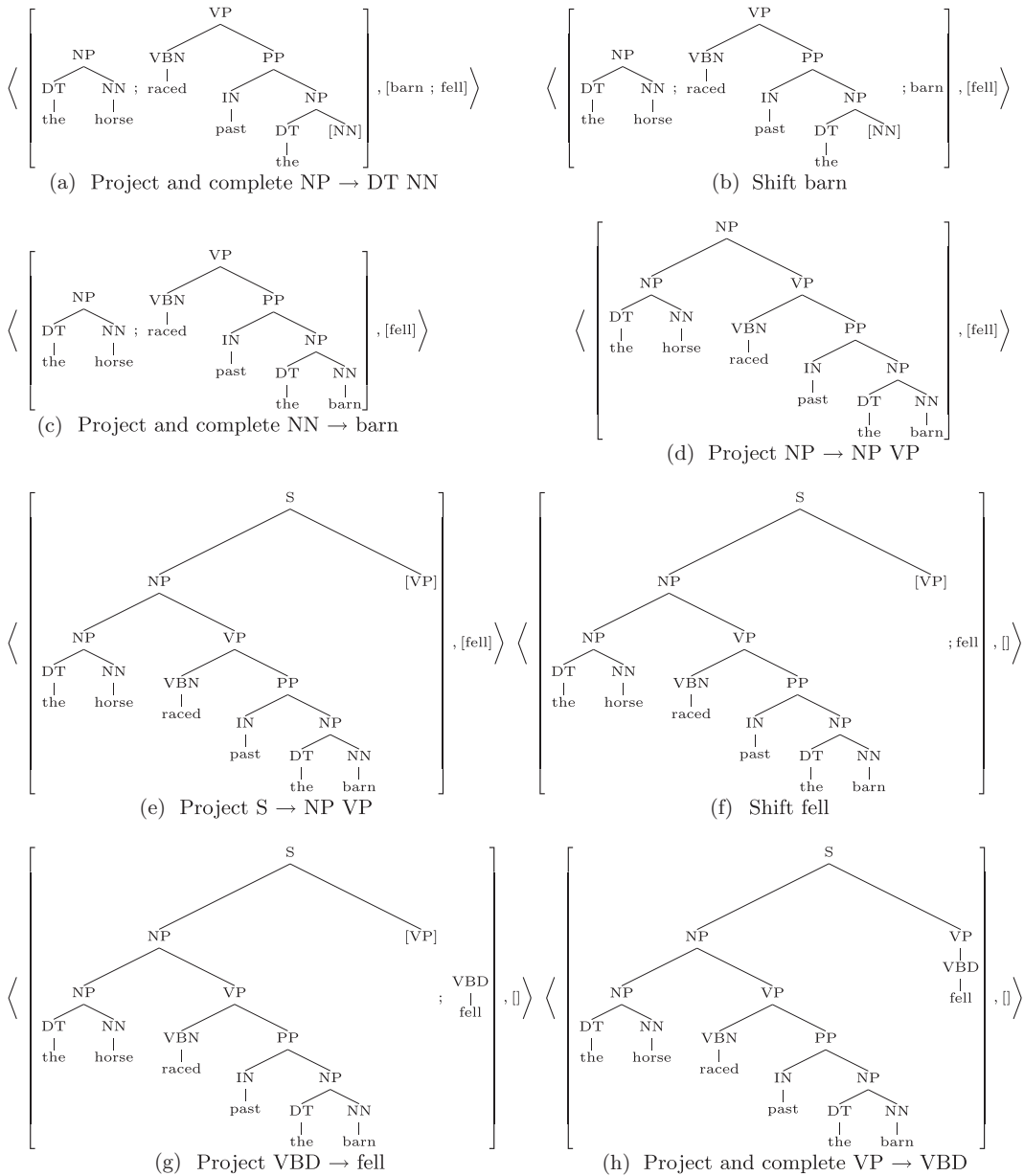


Fig. 6. Illustration of the simple model's mixed parsing strategy.

be implemented by viewing $g(n)$ as the number of GLC steps taken to reach a particular search node. This then prompts the question, how many steps should be expected to finish parsing from search node n ? In other words, how to define h ? Perhaps the most straightforward answer to this question would be to say that estimates about how sentences are going

to end should be based on experience. This idea can be formalized by simulating the action of a given parsing strategy on a corpus sample and recording how far away particular states were from actually finishing. Table 3 collects some of the results of this procedure when applied to the Brown corpus (Kučera & Francis, 1967; Marcus et al., 1993) on the strategy described above in Section 3.1. Each row of Table 3 specifies the contents of a stack, which grows to the left. In stack specifications, categories surrounded with brackets are predictions about upcoming material, while unbracketed categories have been found bottom-up. The table associates a stack with a distance estimate. For instance, in stacks where a determiner (DT) is being sought as part of an NP, which is itself part of a sentence (S), on average those states are 34.478 steps away from finishing an analysis.⁴ This value can be used as the completion-cost estimate in A* search. In other words, one can define $\hat{h}$ explicitly as a lookup table that represents a comprehender’s previous experience with a class of similar states.

The rows of Table 3 describe some of the most frequently experienced intermediate states in the Brown corpus. Others have much lower attestation counts or are not experienced at all. In view of speakers’ infinite creativity, it is inevitable that comprehenders will arrive in novel states to which no past experiences are directly relevant. To model this sort of generalization, the simple model implements a simple state-abstraction scheme. If a state is encountered for which insufficient experience is available, the simple model substitutes experiences with a shorter stack that shares initial elements with the one being considered. It falls back to the longest substack for which enough experience is available. In the simulations to be described in Section 4, “enough” means at least one attestation in the 53,153-sentence Brown corpus. Setting the threshold this low corresponds to the view that some information is better than no information. The particular threshold value is an arbitrary aspect of the simple model’s implementation. The essential idea is that the part/whole rela-

Table 3
Stack configurations attested in the Brown corpus

Stack Contents	Attestations	Average No. of Steps to Goal	SE
VP [S]	55,790	44.936	0.1572
[S]	53,991	10.542	0.0986
NP [S]	43,635	33.092	0.1633
[NP]	38,844	55.791	0.2126
[NP] S [S]	34,415	47.132	0.2122
S [S]	33,578	52.800	0.2195
PP [S]	30,693	34.454	0.1915
[IN] PP [S]	27,272	32.379	0.2031
[DT] NP [S]	22,375	34.478	0.2306
AUX VP [S]	16,447	46.536	0.2863
[VBD] VP [S]	16,224	43.057	0.2826
[VB] VP [S]	13,548	40.404	0.3074
[the] NP [S]	12,507	34.120	0.3046
[NP] NP [S]	12,092	43.821	0.3269
[DT]	10,440	66.452	0.3907

tionship on stack configurations can be used to support generalization across contexts where there is insufficient experience.

3.3. Interpreting the model

The number of states explored during parsing indexes the model's search effort. If the experience-based heuristic is helpful in quickly finding a good syntactic analysis, then fewer states will be explored. If not, then search is less focused and relatively more work is done. As an example of this, note that the simple model finds the globally correct analysis of *the horse raced past the barn fell* after exploring 115 states. By contrast, the globally correct analysis of *the horse raced past the barn* is found after visiting just 43 states. In view of this asymmetry, one says that the model derives the garden-path effect in this case.⁵

This algorithmic-level model characterizes not only how many states are visited but also their content and sequencing relative to one another. For instance, 58 of the 115 states mentioned above treat *the horse* as an unmodified subject. These states entertain the garden-path analysis, and their exploration occurs after the search process first encounters the verb *raced*.

It would be possible to define a detailed linking hypothesis by fitting durational values to classes of GLC transitions. For instance, Lewis and Vasishth (2005) adopt the standard ACT-R assumption that all transitions take 50 ms, plus whatever latency is incurred by memory retrieval. For the models considered in this article, no such fitting has been attempted. In the cases to be considered in following sections, any linking theory that maps steps to approximately equal increments of observed difficulty—such as reading time—would be adequate.

Another issue in the interpretation of the model concerns the “sunk” character of the costs $g(n)$. Why should a searcher take into account the number of steps already taken to reach a particular node, if those costs have already been paid? The answer is that A* minimizes path length, not exploration cost. Rather than locally optimizing for number-of-states visited, A* searches for an analysis that meets the global criteria of being both consistent with the input and economical in GLC steps. The exploration costs that serve as the model's difficulty metric are naturally seen from this perspective as side effects of a comprehension process striving for accuracy, not just speed.

4. Three examples

This section applies Section 3's simple model to three types of human sentence comprehension phenomena. The first phenomenon is the garden-path effect. Many theories, beginning with Frazier's (1979) Garden Path Theory, have insightfully characterized this phenomenon. Its treatment here shows how the new proposal covers an important class of existing observations. It also serves an expository role, showing in detail how the simple model applies to a small, familiar example.

The second phenomenon is one of many counterexamples to the Garden Path Theory and related approaches such as Pritchett's Theta Reanalysis Constraint (1988, p. 545). As an

attachment “paradox,” this example is emblematic of the diverse class of anomalies that threaten any theory of fixed structural preferences (Lewis, 1993, §2.4.2).

The third phenomenon is called local coherence. First characterized by Tabor et al. (2004), it has served as part of an influential argument against processing models that respect the *competence* hypothesis. Konieczny and Müller (2006) find the same sort of effect in German and suggest that it rules out all proposals except those that are dynamical systems in the sense of van Gelder (1998).

The model-fitting exercise carried out in each case is intended not to show a tighter correlation than previous attempts, but rather to show how observed contrasts between experimental conditions follow from a simple model at the algorithmic level. Success with this simple model strengthens the plausibility of the framework as a whole, setting the stage for further development.

4.1. Garden pathing

One of the classic garden paths that Frazier and Clifton (1996) review is shown below in Example 1:

- (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 (Bailey & Ferreira, 2003; Carroll & Slowiaczek, 1991; Speer, Kjølgaard, & Dobroth, 1996). A variety of factors have been implicated in the ambiguity resolution task that Example 1 presents. For instance, lexical properties of the verb *mend* have been shown to affect attachment preferences strongly in these cases (Ford et al., 1982; MacDonald et al., 1994; Mitchell, 1987; Stowe, 1989). In what follows, we put aside such lexical properties in using the simple model to derive the mere possibility of garden pathing. A more detailed picture would undoubtedly emerge from more complex models based on lexicalized grammars.

All of the examples in Section 4 use categories taken from the Penn Treebank (Kučera & Francis, 1967; Marcus et al., 1993). Using the particular grammar specified in Fig. 7, the simple model’s A* search explores the problem space shown in Fig. 8. Circles in such diagrams denote search nodes; each one contains a stack of phrase structures and a list of unparsed words as in Fig. 6. Numbers written inside the circles are timestamps reflecting the order in which informed search visited the node. These pictures detail how informed search arrives at the globally correct analysis in state 43, but not without first exploring the garden-path analysis, states 18–24 (state numbers are shown in Fig. 9). By contrast, the garden path is not explored while parsing the control stimulus 1b. The problem space for this item, Fig. 8(b), shows how informed search proceeds much more directly to the globally correct analysis.

Fig. 9 zooms in states 12 through 25 from Fig. 8(a). The decision to explore the garden path comes after shifting the indefinite determiner *a*. This word is integrated into two

S	→	NP	VP	
S	→	SBAR	S	
S	→	SBAR	,	S
SBAR	→	IN	S	
VP	→	VBG		
VP	→	VBG	NP	
VP	→	VBD	PP	
VP	→	AUX	VP	
VBG	→	mending		
VBD	→	fell		
VBD	→	was		
PP	→	IN	NP	
NP	→	NNP		
NP	→	DT	NN	
NNP	→	Mary		
NN	→	sock		
NN	→	floor		
IN	→	while		
IN	→	on		
DT	→	the		
DT	→	a		

Fig. 7. Grammar for garden pathing in Example 1.

different syntactic contexts in states 18 and 25, respectively. Progress along the globally correct path, which is committed to the intransitive rule $VP \rightarrow VBG$, is rated as more expensive by the experience-based heuristic than returning to the transitive analysis in state 14. Because of this asymmetry in heuristic values, A^* returns to state 14 and does extra work in 1a but not 1b. On the basis of the information provided by the experience-based heuristic, it is rational to explore the garden path in 1a but not in 1b. The branches in subfigure 8(b) reflect subcategorization possibilities common to both items.

The simple model examines 43 search nodes during the analysis of the temporarily ambiguous item 1a. However, it dispatches the unambiguous item 1b after only 38 nodes, despite that sentence having an additional token, the comma. This derives the asymmetry observed by Frazier and colleagues. Although Frazier (1979) accounts for the effect with the Minimal Attachment (MA) heuristic, the present model recasts it as a side effect of optimistic expectations about the way sentences typically end. Rerunning the same simulation with all step-costs set identically to zero demonstrates that this particular prediction is not due to $g(n)$ implementing MA.⁶ Of course, retaining the general formulation with $g(n)$ anticipates increasingly detailed models that incorporate richer notions of analysis quality.

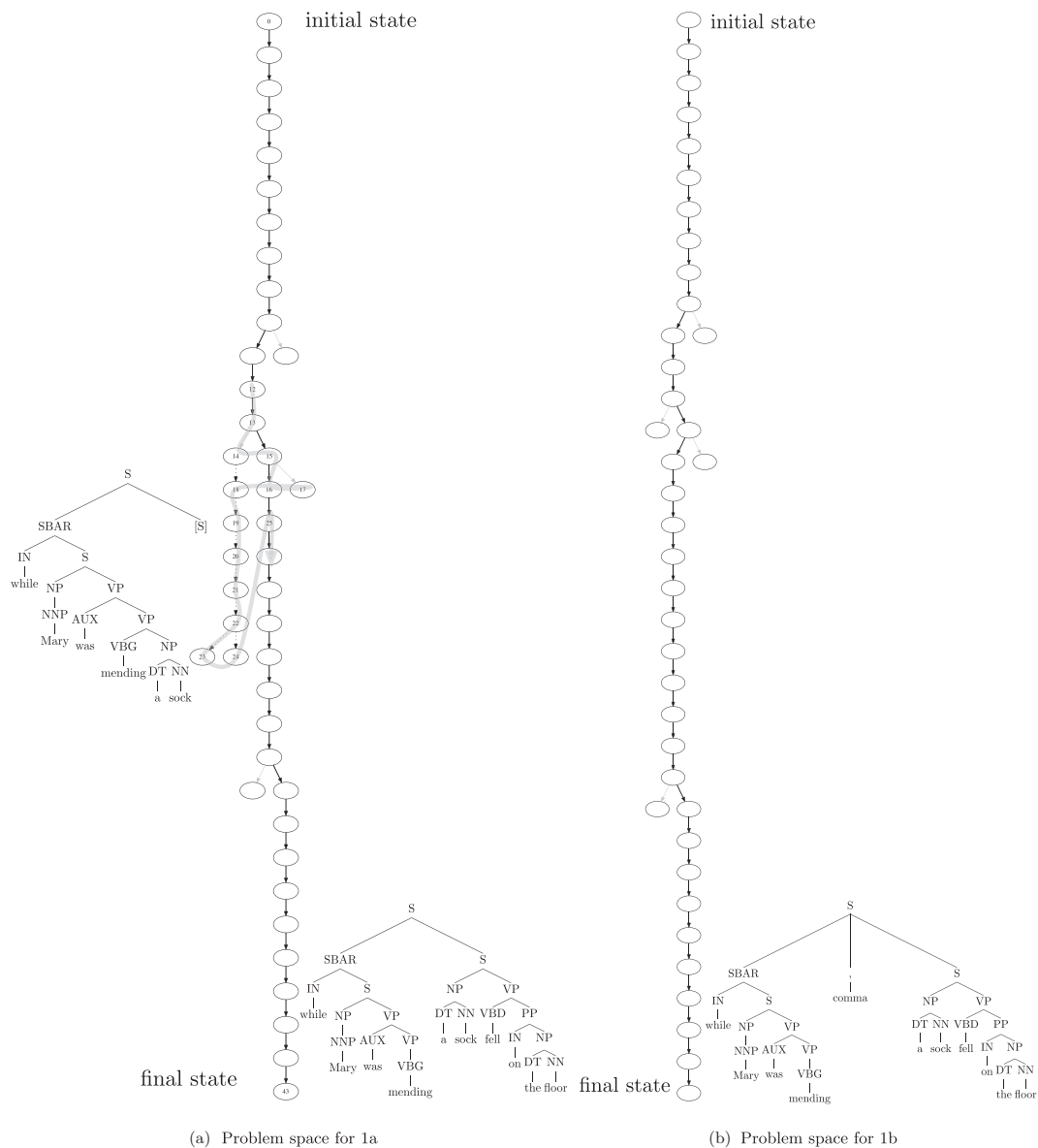


Fig. 8. Search tree with timestamps. Circles denote parser states, and lines denote operator applications. The garden path is indicated with dotted lines. Parser states include pieces of phrase structure licensed by the grammar in Fig. 7.

The generalization behavior of the search heuristic is detailed in Appendix A. This appendix reports, for every lookup, the number of observations that support the $\hat{h}(n)$ estimates. The pattern is the same throughout all of the examples in the article: In parser states with specific lexical words, or deeply nested structure, there is correspondingly more state abstraction and thus, more generalization.

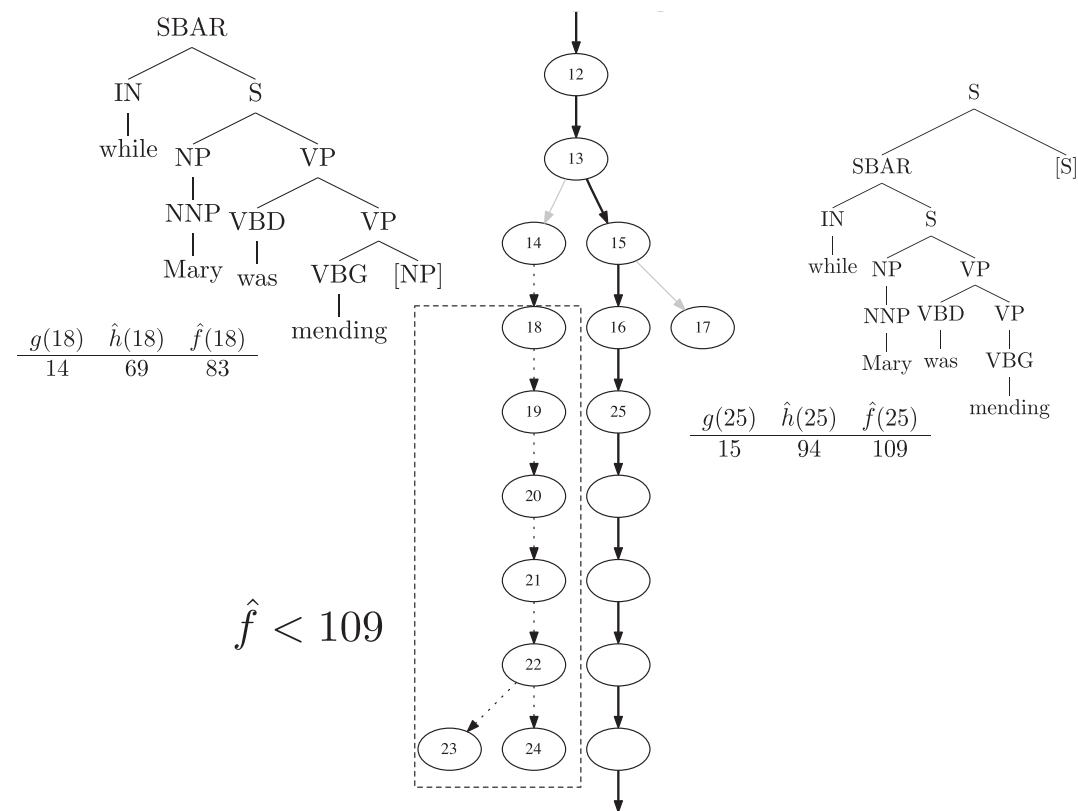


Fig. 9. Subset of search nodes visited during parsing of 1(a).

Another key aspect of the simulation is monotonic structure-building. Models within the framework use parallelism to return to previously unexplored analyses, rather than any kind of repair operator (Grodner, Gibson, Argaman, & Babyonyshev, 2003). Indeed, the simple model's parsing strategy introduced in Section 3.1 and applied in this example does not include any destructive operations.

4.2. Attachment paradox

The simple model's parallelism allows it to overcome paradoxical counterexamples to theories based on fixed attachment preferences such as Right Association or Late Closure (Kimball, 1973). Example 2, adapted from Gibson (1991, p. 22), exemplifies the problem.

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

Example 2 would be categorized as UPA7 in Lewis's (1993) catalog of unproblematic ambiguities. Gibson cites it as evidence against a large class of theories, including the Theta

Reanalysis Constraint (Pritchett, 1992, p. 94). His argument hinges on the different ways that the verb *give* is used in each example. In 2a, *give* appears in the subcategorization frame $\boxed{\text{NP}_{\text{recipient}} \text{NP}_{\text{theme}}}$ whereas in 2b it is attested in the frame $\boxed{\text{NP}_{\text{theme}} \text{PP}_{\text{recipient}}}$. The sentence-final prepositional phrase (PP) is an optional temporal modifier in 2a but a subcategorized element in 2b. An attachment preference like Right Association or Late Closure would eagerly attach *earrings* to *her*; such a preference leads a parser up the garden path in 2a but not 2b. Replacing this heuristic with the opposite principle (perhaps ‘Early Closure’ or ‘Left Association’) would simply swap the predictions such that 2b becomes the predicted garden path instead of 2a. No matter which way fixed preferences are formulated, they derive a prediction of garden pathing on half the items. In fact, neither is particularly hard to understand.⁷

Fig. 10 shows the search spaces explored during the parsing of 2a and 2b, respectively. These states explore analyses defined by the grammar in Fig. 11; the final states in each case reflect the globally correct analyses shown in Fig. 12. Consistent with the Treebank annotation guidelines, the grammar specifies that the word *her* may either be possessive (PRP\$) or nonpossessive (PRP). The problem spaces show how the simple model copes with this temporary ambiguity. In these diagrams, filled or darkened circles indicate states in which the ‘other’ interpretation of *her* is explored. That is, the highlighted states have a grammatical interpretation of 2a that fits the globally correct reading of 2b and vice versa. Categorizing these search spaces according to whether or not they allocate effort finding the other analysis identifies the parsing work shared across 2a and 2b.

Much of the work turns out to be shared. The experience-based heuristic leads the simple model to interleave the garden-path attachment and the globally correct attachment. The result is a search process that is strictly committed to neither analysis. In 2a, fifteen states represent the globally incorrect attachment of *her*. In 2b, fourteen 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.

4.3. Local coherence

Having considered garden pathing and counterexamples to it, this section turns to a phenomenon that has come to light more recently called local coherence. It was first characterized in English by Tabor et al. (2004) and in German by Konieczny (2005); Konieczny and Müller (2006, 2007). This phenomenon is significant because it figures in influential arguments for dynamical, rather than computational, models of cognition. The dynamical approach (van Gelder, 1998; Port, 2003) sets itself against certain received positions in classical cognitive science such as:

1. A commitment to discrete and context-free symbols ...
2. The view of rules as operators ...
3. The static nature of representations ...
4. The building metaphor ...

(Elman, 1995, pp. 198–199)

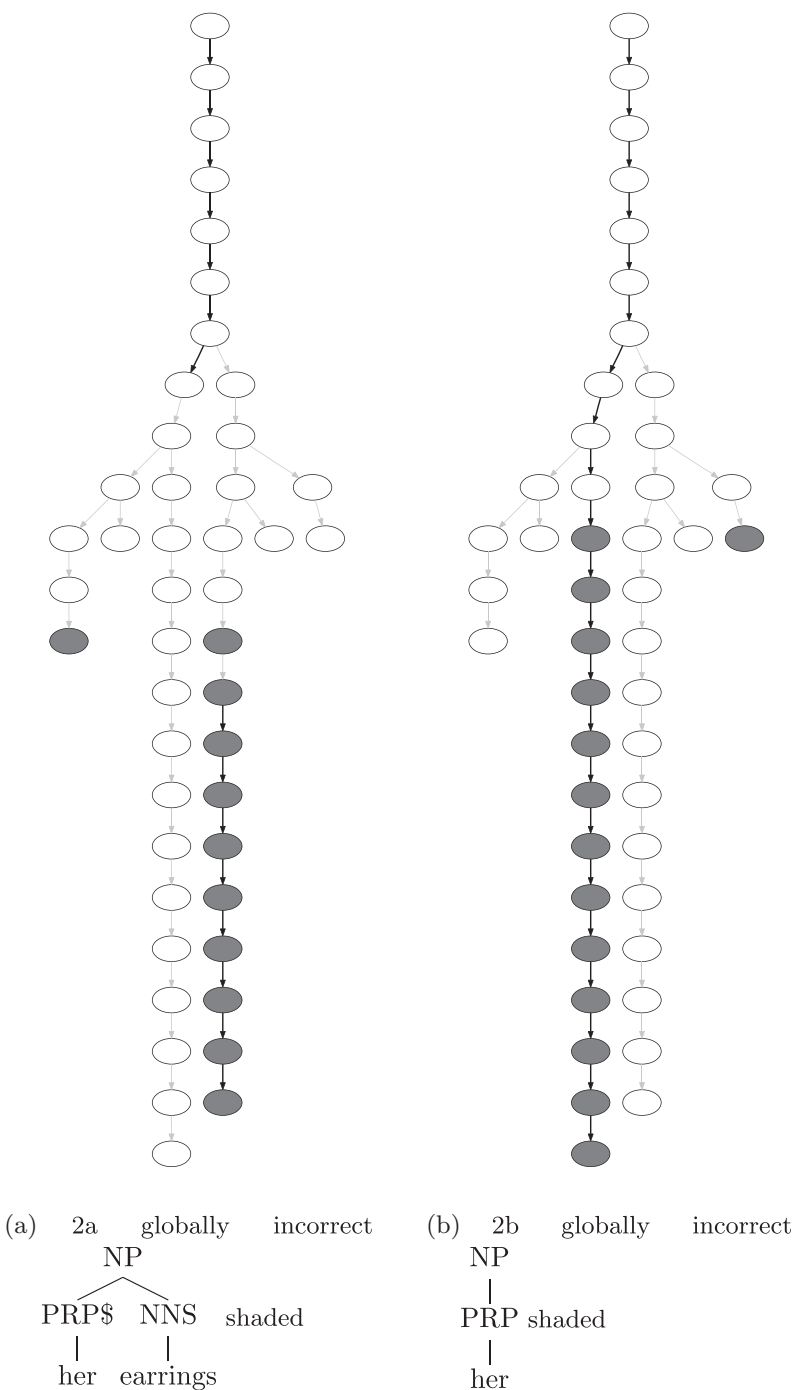


Fig. 10. It is rational to do a little of both. Circles denote search nodes; darkened circles indicate states in which the “other” interpretation is being explored.

S	→	NP	VP		
VP	→	VBD	NP	PP	
VP	→	VBD	NP	NP	PP
VBD	→	gave			
TO	→	to			
PRP\$	→	her			
PRP	→	her			
PRP	→	I			
PP	→	TO	NP		
PP	→	IN	NP		
NP	→	PRP\$	NNS		
NP	→	PRP\$	NN		
NP	→	PRP			
NP	→	NNS			
NP	→	DT	NN		
NNS	→	earrings			
NN	→	girl			
NN	→	birthday			
IN	→	on			
DT	→	another			

Fig. 11. Grammar for the unproblematic ambiguity in Example 2.

These assumptions lie at the heart of the broad framework introduced in Section 2. This section argues that the local coherence data are in fact compatible with the broad framework. The argument proceeds by example, taking up first the English finding, then the German finding. This modeling also improves upon previous work based on Markov models. This class of theories is identified as LCA2 in Tabor et al. (2004) and studied further in Bicknell, Levy, and Demberg (2009). Such devices do not possess any sort of top-down, predictive component. As Gibson (2006) writes, “The top-down component...needs to be worked out in more detail in order to provide an analytic model” (p. 365). The simple model, with its partially top-down parsing strategy, offers a way to understand such a hypothesized component of the human sentence processing mechanism.

4.3.1 English

The term “local coherence” itself refers to a substring that can be given an independent syntactic analysis that does not fit into the globally correct analysis. For instance, in Example 3a, the substring *the player tossed a frisbee* could be analyzed as a sentence as shown in Fig. 13. In the context of Example 3 such an analysis is grammatically impossible: The preposition *at* does not take a sentence as a syntactic complement in English.

- (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

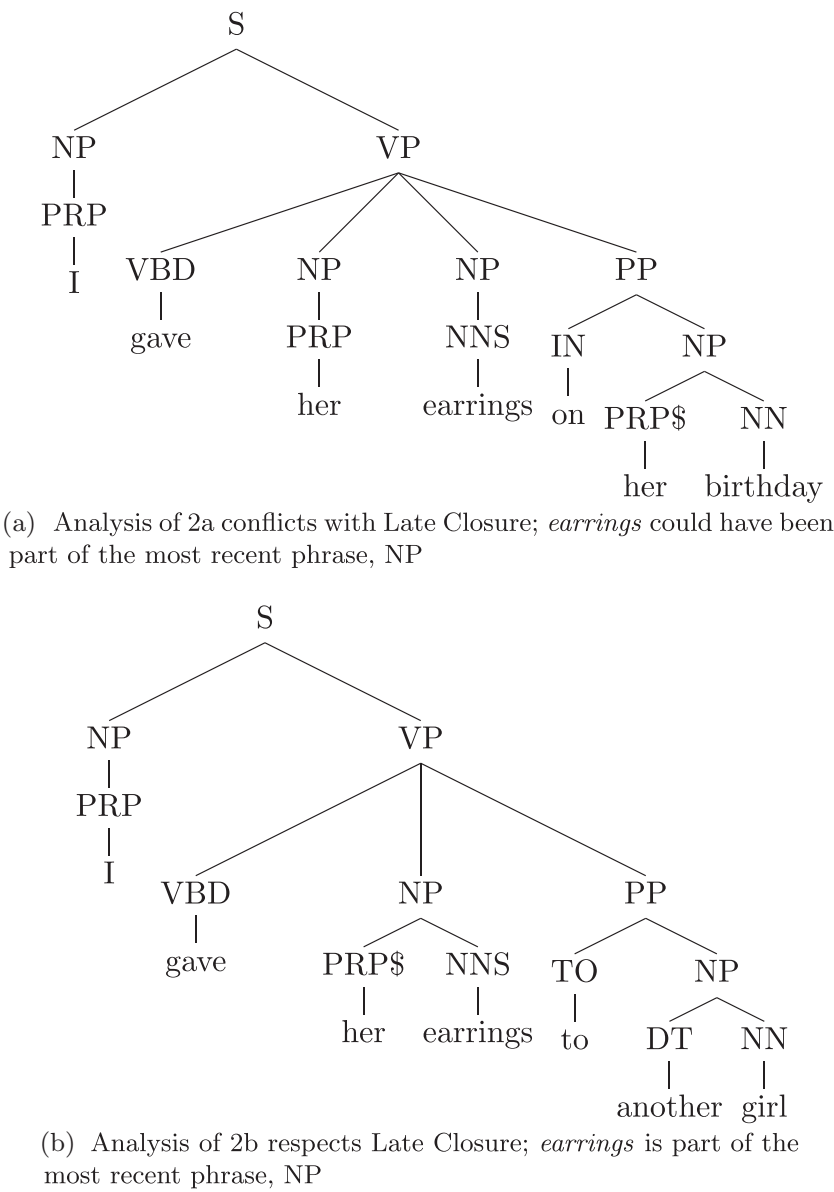


Fig. 12. Attachment paradox for fixed heuristics.

Local coherence seems different from garden pathing. In classic temporary ambiguities like Example 1, some analysis of the ambiguous region turns out to be grammatically inconsistent with material that eventually arrives on its right. By contrast, a locally coherent substring may be inconsistent as well with material on its left. Tabor and colleagues demonstrate that the presence or absence of these left-inconsistent analyses in English has a measurable effect on reading time. Their experiment manipulates two binary factors:

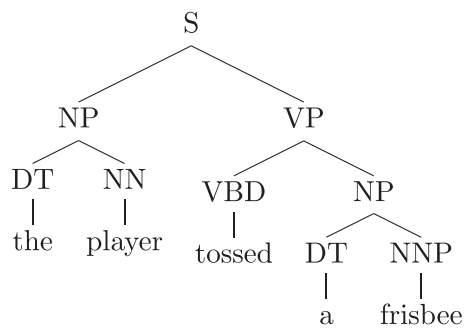


Fig. 13. Locally coherent analysis not compatible with global structure of 3a.

$\pm$ reduction, which amounts to the presence or absence of *who was*, and $\pm$ ambiguous. A stimulus that is +ambiguous employs a verb like *tossed* that is homophonous in active voice and passive voice, whereas embedded verbs in the $-$ ambiguous stimuli always have a spelling that is unique to the passive. Table 4 summarizes this 2×2 design, highlighting the cell in which stimuli contain the locally coherent substring. Global syntactic analyses for these items are shown in Fig. 14.

Tabor et al. (2004) analyze self-paced reading times in a region comprising the embedded verb (e.g., *tossed* or *thrown*) and the following three words. In this region, they observe an interaction between the factors $\pm$ ambiguous and $\pm$ reduced such that the well-known slowdown associated with reduced relative clauses is exacerbated when the embedded verb is ambiguous. This observed superadditivity suggests that the embedded verb is indeed considered as a main verb for its own sentence, a sentence in which *the player* functions as a subject. A key precondition for any modeling based on this intuition is that the model be able to explore locally coherent states. In some parser designs, a prediction table or left-corner table would prevent such exploration. The table would encode global knowledge of the fact that sentences cannot be complements of prepositions (Earley, 1970; Stolcke, 1995). Nothing in the simple model requires such a table. In fact, the simple model allows for the possibility that such unfinishable states could be explored.

Parsing runs with the simple model confirm that locally coherent states are, in fact, explored in stimuli like the ones Tabor and colleagues used. To ensure that none of the predictions come down to out-of-vocabulary words, the word *frisbee* is replaced by *ball* in the input to the model.⁸ Fig. 15 illustrates this additional searching by depicting the crucial sector of search space in items like Example 3a and 3c. In the region that Tabor and

Table 4
Design of Tabor et al. (2004) Experimental 1. Locally coherent cell is shaded

	+ambiguous	–ambiguous
+reduced	3a	3b
–reduced	3c	3d

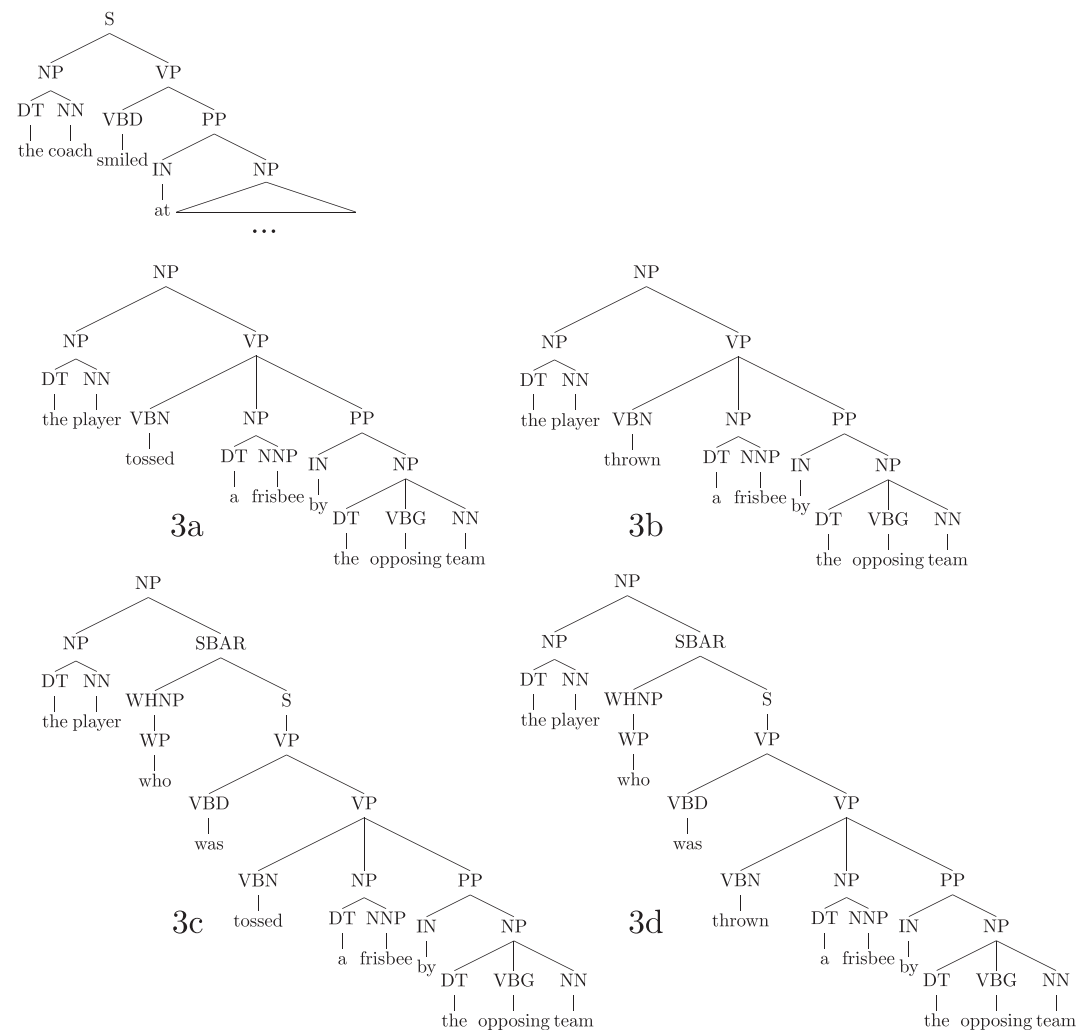
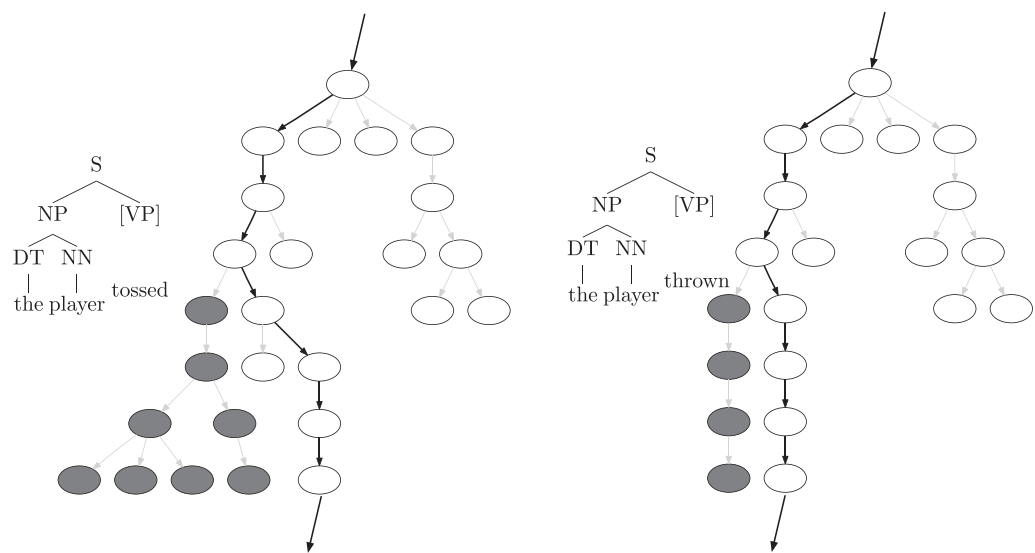


Fig. 14. Syntactic structures for the Tabor et al. (2004) stimuli in Penn Treebank style.

colleagues analyzed, there is an ambiguity penalty of five states in the ambiguous *tossed* case compared to a corresponding penalty of only three states in the *thrown* case. This amplified effect of reduction in the ambiguous case follows jointly from the parsing strategy, the grammar and the heuristic.

To see how it follows, consider the simple model's progress immediately after the words *the player*. The globally correct analysis in all conditions demands a postmodifying relative clause. Because the simple model's parsing strategy does not predict postmodifiers, both the modifi-ee NP and its modifi-er must be built bottom-up. By the time this has been accomplished, three pieces of structure have been built as shown in Fig. 16. There are 18 cases in the Brown corpus where a state like this one is attested. On average, these states are 76 steps away from finishing. This low level of "optimism" about the completion cost $\hat{h}$ leads A* to



(a) Parsing 3a visits more locally-coherent states (b) Parsing 3b visits fewer locally-coherent states

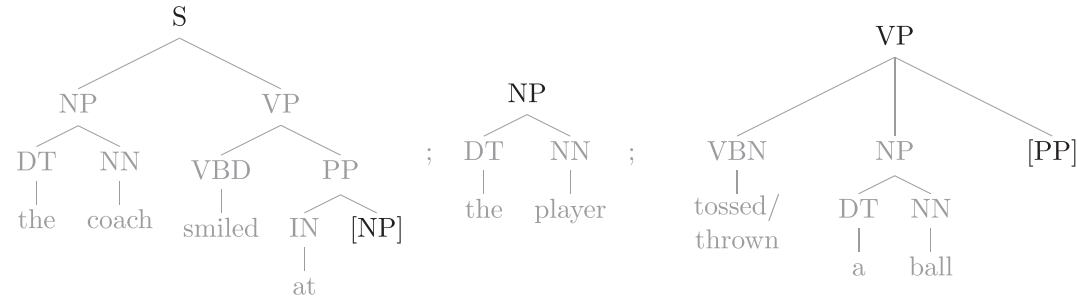
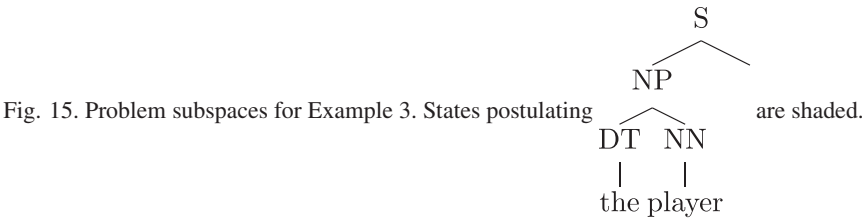


Fig. 16. Intermediate state that motivates reconsideration: [PP] VP NP [NP] S. Note that while stack specifications in this paper grow from right to left, the phrase structures above are displayed from left to right.

examine the locally coherent alternatives. These are the states that result from projecting a sentence node over *the player* as opposed to leaving it on the stack to await combination with a postmodifier. These analyses have yet to get as far as the word *ball* and thus, it is rational to develop them so that their costs can be compared with the analyses shown in Fig. 16. This reconsideration exercises all of the possibilities provided in the grammar, Fig. 17. As there are more ways to use *toss*—for instance, more subcategorization

WP	→	who		
WHNP	→	WP		
VP	→	VBN	NP	PP
VP	→	VBD	VP	
VP	→	VBD	PP	
VP	→	VBD	NP	
VBN	→	tossed		
VBN	→	thrown		
VBG	→	opposing		
VBD	→	was		
VBD	→	tossed		
VBD	→	smiled		
TOP	→	S		
S	→	VP		
S	→	NP	VP	
SBAR	→	WHNP	S	
PP	→	IN	NP	
NP	→	NP	VP	
NP	→	NP	SBAR	
NP	→	DT	VBG	NN
NP	→	DT	NNP	
NP	→	DT	NN	
NNP	→	frisbee		
NN	→	team		
NN	→	player		
NN	→	coach		
IN	→	by		
IN	→	at		
DT	→	the		
DT	→	a		

Fig. 17. Penn Treebank-inspired grammar for the English local coherence cases.

possibilities—more states are explored in 3a than 3c. All of these eventually come up against a wall of very high cost, which leads A* back to the globally correct pathway.

The simple model correctly derives the observed superadditive interaction of $\pm$ reduction and $\pm$ ambiguity because the $\hat{h}$ values rationally motivate consideration of locally coherent states. The prediction is independent of $\hat{h}$'s generalization threshold, but it does depend on the search procedure being free to explore locally coherent states. The empirical distance heuristic could have led the parser away from these states, but it in fact does not. This outcome spotlights the trade-off between grammatical deduction and graded heuristic guidance. It may be that alternative behavioral tasks and individual differences can be modeled in terms of the trade-off between these two theoretical devices. In-depth exploration of these possibilities should be informed by future work, both empirical and theoretical.

4.3.2 German

Konieczny and Müller (2006) observe a similar kind of local coherence in German using stimuli like those in Example 4.

- (4) a. Die Tatsache, dass die Astronautin überrascht den Außerirdischen
 the fact, that the astronaut_{fem} surprisedly/surprises the alien
 vom Mars entdeckte, erregte Aufsehen.
 from Mars discovered, caused a sensation
- b. Die Tatsache, dass die Astronautin ungläubig den Außerirdischen vom
 the fact, that the astronaut_{fem} perplexedly the alien from
 Mars entdeckte, erregte Aufsehen.
 Mars discovered, caused a sensation

Example 4a uses a form of *überraschen*, which is ambiguous between an adverb and a present-tense inflected main verb. This latter possibility allows for the locally coherent interpretation shown in Fig. 18. By contrast, the control item 4b has a very similar meaning but lacks the ambiguity that would lead to a local coherence.

If no global filtering is applied, A* search arrives at the globally correct analysis of 4a, shown in Fig. 19(b), after visiting 1,390 states. It visits only 310 states on the way to the analysis of 4b in Fig. 19(b). In the locally coherent example, 158 of the explored states include an analysis of [_{NP} die Astronautin] as a sentence subject. During parsing of the control item, only four such states are visited. This asymmetric prediction is consistent with the local coherence effect observed by Konieczny and Müller (2006). In adapting the simple model to this example, the grammar in Fig. 20 is used in conjunction with distance-to-completion metrics read-off from the Tiger Treebank (Brants et al., 2004).⁹

4.4. Discussion of the three examples

The preceding subsections have sought to lay bare the functioning of the simple model using small grammars that lead to relatively small problem spaces. Scaling the model up to

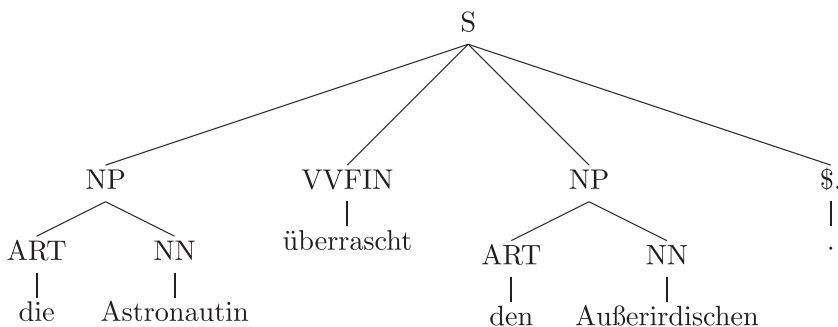
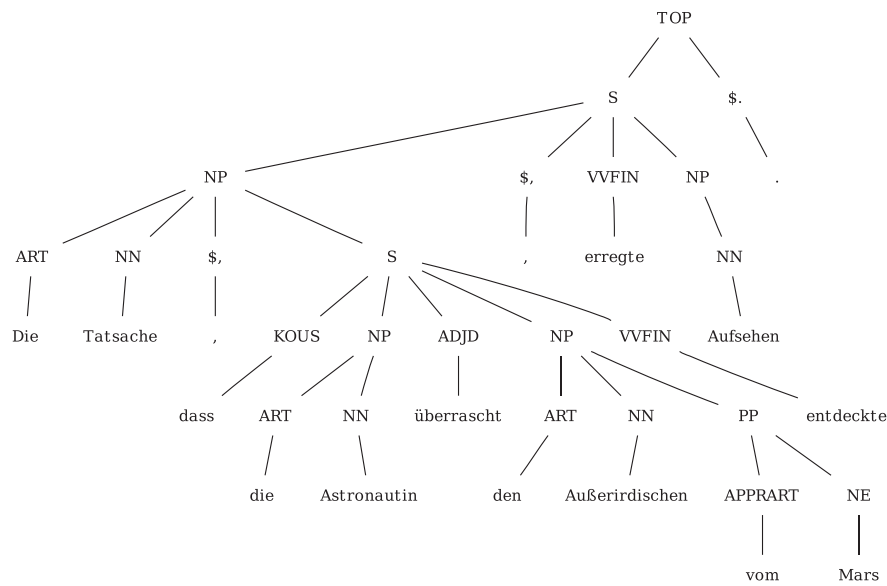
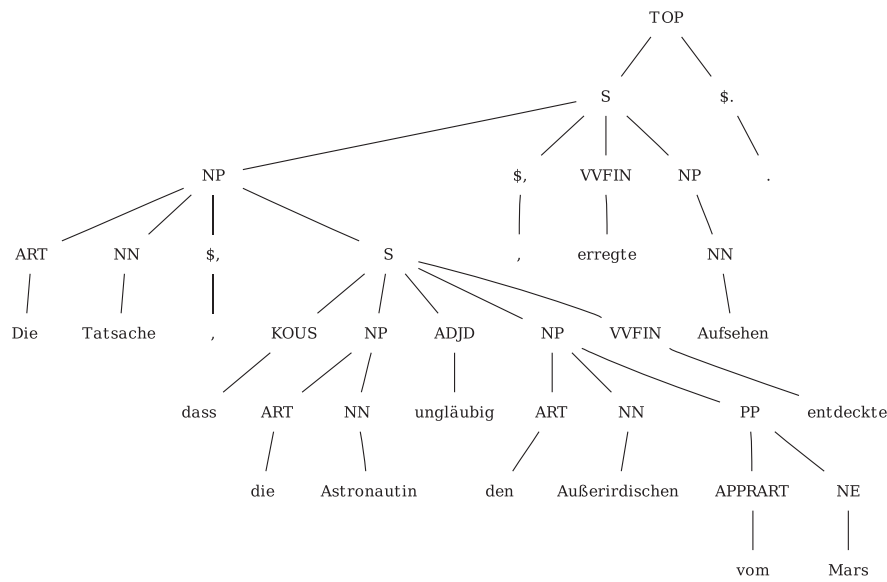


Fig. 18. Local coherence in 4a.



(a) Syntactic analysis for example 4a



(b) Syntactic analysis for example 4b

Fig. 19. NEGRA format analyses for the the stimuli in 4.

S	→	NP \$, VVFIN NP
S	→	NP VVFIN NP
S	→	KOUS NP ADJD NP VVFIN
NP	→	NN
NP	→	ART NN
NP	→	ART NN \$, S
NP	→	ART NN PP
PP	→	APPRART NE
VVFIN	→	überrascht
VVFIN	→	erregte
VVFIN	→	entdeckte
NN	→	Tatsache
NN	→	Außerirdischen
NN	→	Aufsehen
NN	→	Astronautin
NE	→	Mars
KOUS	→	dass
ART	→	die
ART	→	den
ART	→	Die
APPRART	→	vom
ADJD	→	überrascht
ADJD	→	ungläubig

Fig. 20. Tiger-inspired grammar for German local coherence cases.

larger grammars enlarges the problem spaces with new options to pursue. The three English examples have been examined with a larger “merged” grammar that combines all of the English rules mentioned in any of the examples, and search within this grammar generally mimics the relationships found with the individual grammars. One exception to this generalization is PP attachment. The detailed balance between subcategorization frames in Section 4.2’s attachment paradox is sensitive to number and type of PP attachment options. This sensitivity does not impugn the claim that the simple model can interleave its consideration of more than one analysis. Rather, it points to the many information sources—such as degree-of-plausibility—that are excluded from the simple model’s empirical heuristic. While this heuristic provides perhaps the most direct formulation of the chose-based-on-experience idea, its status as a lookup table renders it difficult to draw generalizations about the particular type or types of knowledge that it encodes. To gain a deeper understanding of what this heuristic actually does, the next section considers a related heuristic which encodes some of the same generalizations.

5. An entropy-reducing mechanism

5.1. The heuristic’s knowledge of sought categories

The simple model’s completion cost estimate treats parser states as sequences of category symbols. What can be said about the contribution of individual symbols within these sequences? Noting the GLC distinction between “sought” and “found” categories, one

might imagine that the “sought” categories would be informative about the distance to completion. After all, if a full parse is eventually to be found, then full parses for each of the remaining sought categories will necessarily need to be found. They serve as promises about work that needs to be completed in successor states. If the completion cost estimates read off from a corpus sample indeed reflect these promises, then there should be a correlation between sought categories at a search node n and the estimated distance-left-to-goal $\hat{h}(n)$. Parser states with more sought categories should, in general, be farther from completion if $\hat{h}(n)$ reflects top-down promises. On the other hand, different categories promise different amounts of work. A noun phrase, for instance, is much more likely to have a long derivation than a preposition. Such likelihoods can be formalized in terms of expected derivation length, shown below in Eq. 2, adapted from Wetherell (1980). Wetherell’s notation presupposes a numbering on symbols in a grammar: S might be symbol #1, SBAR would be symbol #2 and so on. The expected derivation length is a vector whose components are the number of rewritings required during an average derivation of the i th nonterminal. The matrix $\mathcal{A}$ is the stochastic expectation matrix for one-step derivations.¹⁰ Its entries $\mathcal{A}_{ij}$ contain the sum of probabilities that the i th grammar symbol is rewritten as the j th symbol. The limit of matrix powers, $\mathcal{A}^\infty$ combines expectations for derivations of all lengths. The identity matrix $\mathbf{I}$ plays a role formally analogous to the number 1 in the closed-form solution of a geometric series of scalars.

$$\begin{aligned} \text{expected derivation length (EDL)} &= \mathcal{A}^\infty \times \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix} \\ \mathcal{A}^\infty &= \sum_{i=0}^{\infty} \mathcal{A}^i = \frac{\mathbf{I}}{\mathbf{I} - \mathcal{A}} = (\mathbf{I} - \mathcal{A})^{-1} \end{aligned} \quad (2)$$

Summing the expected derivation lengths for all sought categories in a state, and restricting consideration to states for which there are sufficient attestations in the Treebank (say, more than 150) a correlation of about 0.4 emerges between the amount of work explicitly promised in a parser state and the empirical estimate of completion distance. Fig. 21 shows this correlation.

5.2. Relationship to entropy

The correlation plotted in Fig. 21 also obtains between the mean-steps-to-completion and the average uncertainty or *entropy* of sought categories in a state. This is perhaps unsurprising given that the entropy of a nonterminal is just the same matrix $\mathcal{A}^\infty$ multiplied on the right by a vector of single-rewrite entropy values. Using the notation Prob for the probability of a rule in a probabilistic context-free grammar (PCFG) the entropy of a single-rule rewriting event in a derivation is given below in Eq. 3. The value h_{XP} quantifies the degree to which the conditional distribution on possible right-hand sides of XP rules is either peaked or flat.

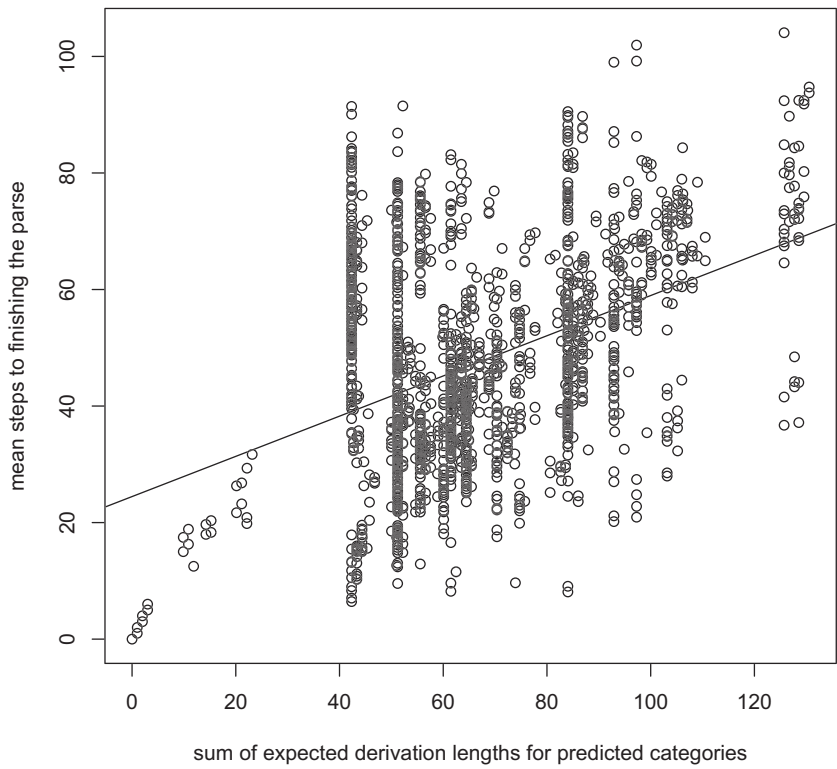


Fig. 21. Correlation between summed expected derivation lengths of categories sought in a parser state and mean number of steps to completion. $r = .4, p < .0001$.

$$h_{XP} = - \sum_{r \in \text{rules of the form } XP \rightarrow \dots} \text{Prob}(r) \log_2 \text{Prob}(r) \tag{3}$$

The h_{XP} values summarize the degree of uncertainty about one-step derivations in a PCFG. Combining them with $\mathcal{A}^\infty$ yields average uncertainty values about any derivation started by the given nonterminal. Equation 4 highlights the formal similarity between the entropy of a nonterminal, and its average derivation length.

$$\text{entropy of nonterminal} = \mathcal{A}^\infty \times \begin{bmatrix} h_S \\ h_{S\text{BAR}} \\ h_{NP} \\ h_{N\text{BAR}} \\ h_{VP} \\ \vdots \end{bmatrix} \tag{4}$$

5.3. Entropophobic heuristics and entropy reduction

Applying EDL or entropy of sought categories as a heuristic in A* search is straightforward—we term this the “refined” model in what follows. Both lead to an avoidance of categories that might be big or complex. If these categories are required in a globally correct analysis, such an “entropophobic” heuristic will explore alternative pathways first. More precisely, if a node n is on the globally correct path, then all alternative nodes with heuristic values $\hat{f}(n') < \hat{f}(n)$ will be searched. This set of “step-sister” nodes n' must all be disconfirmed. If this set is large, the overall transition of the search from the way it was before n to after n will be difficult. From the perspective of the Entropy Reduction Hypothesis (Hale, 2003, 2006; Wilson & Carroll, 1954), the transition from a large set of alternatives to a single alternative also represents work. The following section identifies the relevant sets in drawing a connection between the ERH and the refined model in a specific example.

5.4. Easy versus hard garden paths

This section applies the refined model in a rederivation of the difficulty asymmetry between two garden-path sentences. This asymmetry follows from the Entropy Reduction Hypothesis (Hale, 2003). However, rederiving it using the refined model not only demonstrates how the model applies across sentences but also highlights the relationship of the current proposal to this particular, preexisting rational analysis of parsing.

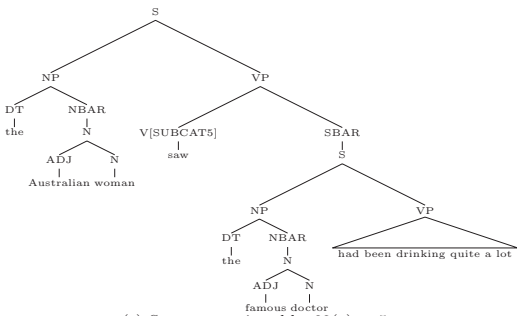
The garden pathing asymmetry under consideration is a relationship between the NP/Z ambiguity discussed above in Section 4.1 and another, milder ambiguity known as NP/S. The NP/S ambiguity, laid out with its control stimulus in 5, comes about when a verb like *saw* can take either an NP complement or a sentence (S) complement. Sturt et al. (1999) contrasted the NP/S with the version of the NP/Z temporary ambiguity shown below in 6. The naming convention reflects the fact that *visited* could either take an NP complement or no complement at all in its intransitive frame; this latter possibility is the zero (Z) case. The empirical finding is that the NP/Z garden-path effect measured in 6 is larger than the NP/S garden-path effect in 5 (Sturt et al., 1999).

- (5) a. the Australian woman saw the famous doctor had been drinking quite a lot
- b. the Australian woman saw that the famous doctor had been drinking quite a lot
- (6) a. before the woman visited the famous doctor had been drinking quite a lot
- b. before the woman visited , the famous doctor had been drinking quite a lot

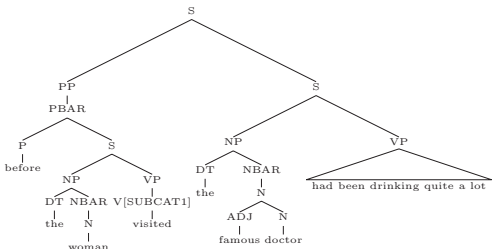
Note that the refined model’s heuristic deals in average derivation length or average uncertainty of particular categories. These values can be computed directly from a weighted grammar. Fig. 22(a) lists weighted grammar rules, inspired by Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, & Sag, 1985), that cover 5 and 6. This grammar extends the one presented in Hale (2003, fig. 12 on p. 115) with additional rules required for the analysis of the control conditions 5b and 6b. In the spirit of Anderson’s (1990) second step,

weight	rule		nonterminal	EDL
75	S	→ NP VP	S	29
13	S	→ PP , S	SBAR	30
12	S	→ PP S	NP	5
1	SBAR	→ S	NBAR	3
1	SBAR	→ that S	VP	12
1	NP	→ DT NBAR	VBAR[PRP,COP]	5
1	NBAR	→ N	VBAR[PRP]	3
1	VP	→ V[SUBCAT2] NP	PP	32
1	VP	→ V[SUBCAT1]	PBAR	31
1	VP	→ V[SUBCAT5] SBAR	P	1
1	VP	→ V[SUBCAT4,ASP] VBAR[PRP,COP]	ADVP	1
1	VBAR[PRP,COP]	→ V[SUBCAT7,PRP,COP] VBAR[PRP]	V[SUBCAT4,ASP]	1
1	VBAR[PRP]	→ V[SUBCAT2,PRP] ADVP	V[SUBCAT7,PRP,COP]	1
1	PP	→ PBAR	V[SUBCAT2,PRP]	1
1	PBAR	→ P S	V[SUBCAT2]	1
1	P	→ before	V[SUBCAT1]	1
1	ADVP	→ quite a lot	V[SUBCAT5]	1
1	V[SUBCAT4,ASP]	→ had	DT	1
1	V[SUBCAT7,PRP,COP]	→ been	ADJ	1
1	V[SUBCAT2,PRP]	→ drinking	N	2
1	V[SUBCAT2]	→ visited		
1	V[SUBCAT2]	→ saw		
1	V[SUBCAT1]	→ visited		
1	V[SUBCAT5]	→ saw		
1	DT	→ the		
1	ADJ	→ famous		
1	ADJ	→ Australian		
1	N	→ ADJ N		
1	N	→ woman		
1	N	→ doctor		

(a) Grammar for NP/S and NP/Z ambiguities extended from Hale (2003)



(c) Structure assigned by 22(a) to 5a



(d) Structure assigned by 22(a) to 6a

Fig. 22. Linguistic properties of the Sturt et al. (1999) stimuli.

one might say that this grammar constitutes a formal account of a speech community that does not front their prepositional phrases very often. The grammar specifies that 75% of the time sentences appear without fronted PPs. The other 25% of the time they are fronted, and the comma is optional.¹¹ Uniform weights on the other rules express a lack of commitment to any special construction frequencies. Note that the grammar describes its weighted language in a way that is independent of any parsing strategy.

Table 5 shows the effort profile of the refined model across these four items. The refined model correctly derives the greater severity of NP/Z as compared to NP/S. In

Table 5
Count of states examined during parsing the Sturt et al. (1999) stimuli

	Experimental	Control
NP/S	48 (5a)	40 (5b)
NP/Z	53 (6a)	42 (6b)

the easier NP/S case (Fig. 23), the globally correct path requires projecting an SBAR complement of *saw*. This happens at state 22 whose $g(22)$ cost so far is 13 steps. As shown in Fig. 22(b), the SBAR category has an average derivation length of 30 steps. Adding this completion cost estimate to the number of steps actually taken so far gives the total heuristic value, $43 = 13 + 30$. To get back on to this globally correct path, A* must disconfirm nine accessible states whose heuristic value is less than 43. The string-level progress of all these states remains on or near the verb *saw*. Metaphorically, this sequence of events might be described as the parser equivocating because it does not want to project a new clause.

In the harder NP/Z case (Fig. 24) the globally correct path requires passage through a state with outstanding predictions for [S], [NBAR], and [N]. At state 37, whose $g(37)$ value is 22, this adds up to $56 = f(37) = 22 + (29 + 3 + 2)$. Shifting *doctor* does not directly fulfill any of these predictions and A* instead returns to the developing garden-path analysis. This analysis entails roughly the same cost in terms of sought categories but has crucially not advanced as far as the globally correct analysis. Metaphorically, this sequence of events might be described as the parser synchronizing two lines of analysis because they both are about as good. Resolving this indecision involves contracting the step-sister set of $\hat{f} < 56$ nodes from thirteen down to one. The fact that this result does not follow when all step costs are set identically to zero confirms the essential role of $g(n)$.

Although the ERH deals in word-to-word transitions, models within the broad framework deal in transitions between states. Each analysis pathway is extended one step at a time. These steps are sequential with respect to a particular syntactic analysis. But their exploration may not occur sequentially in time. If there are other pathways with better heuristic values, they need to be disconfirmed before search can proceed in a given direction. The general connection between the ERH and the broad framework is that the disconfirmation of these alternatives takes more time when this set is large.

6. Related work

The most important prior work on rational analysis in human parsing is by Chater et al. (1998; see also Crocker, 2005). Section 5.4 treats the NP/S ambiguity they discuss. Beyond this one example, Chater and colleagues argue more generally that successful parsing shies away from choices that are hard to recover from. When parsing cannot return to a previous choice point, this is called “crashing.”

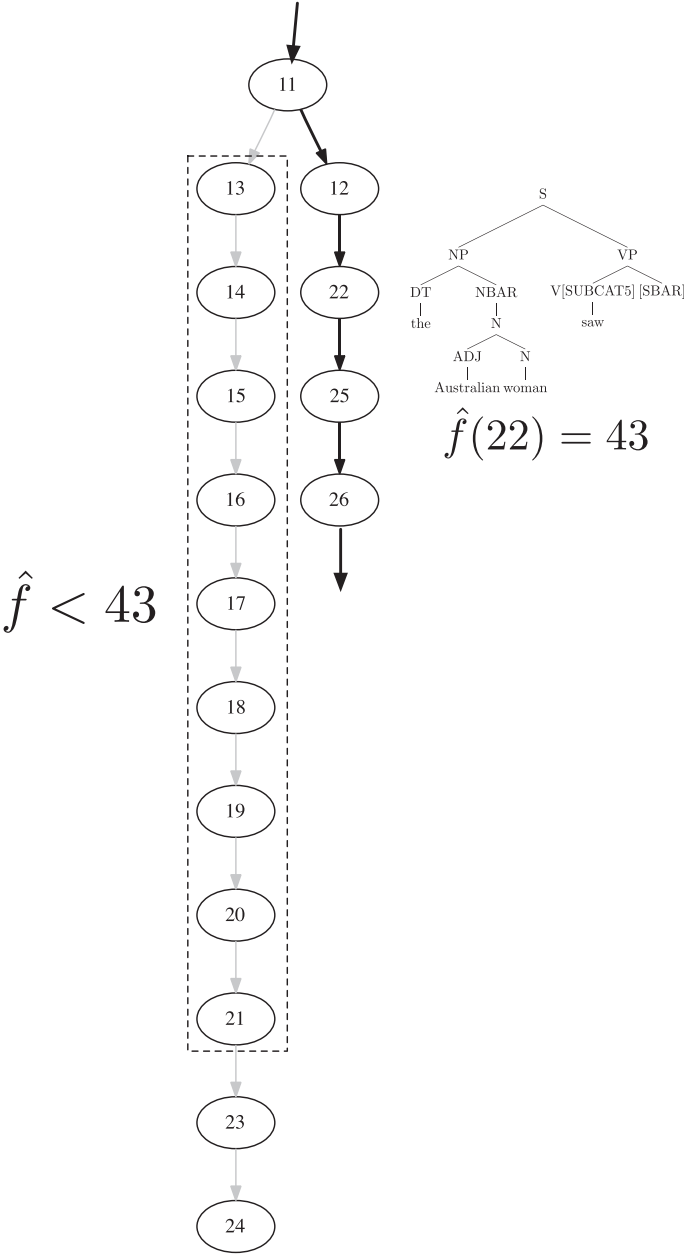


Fig. 23. Problem subspace for the “easy” NP/S garden-path sentence.

We assume that the parser chooses the order in which to consider alternative hypotheses at a point of local ambiguity to minimize the overall probability of a crash in parsing a sentence. (p. 456)

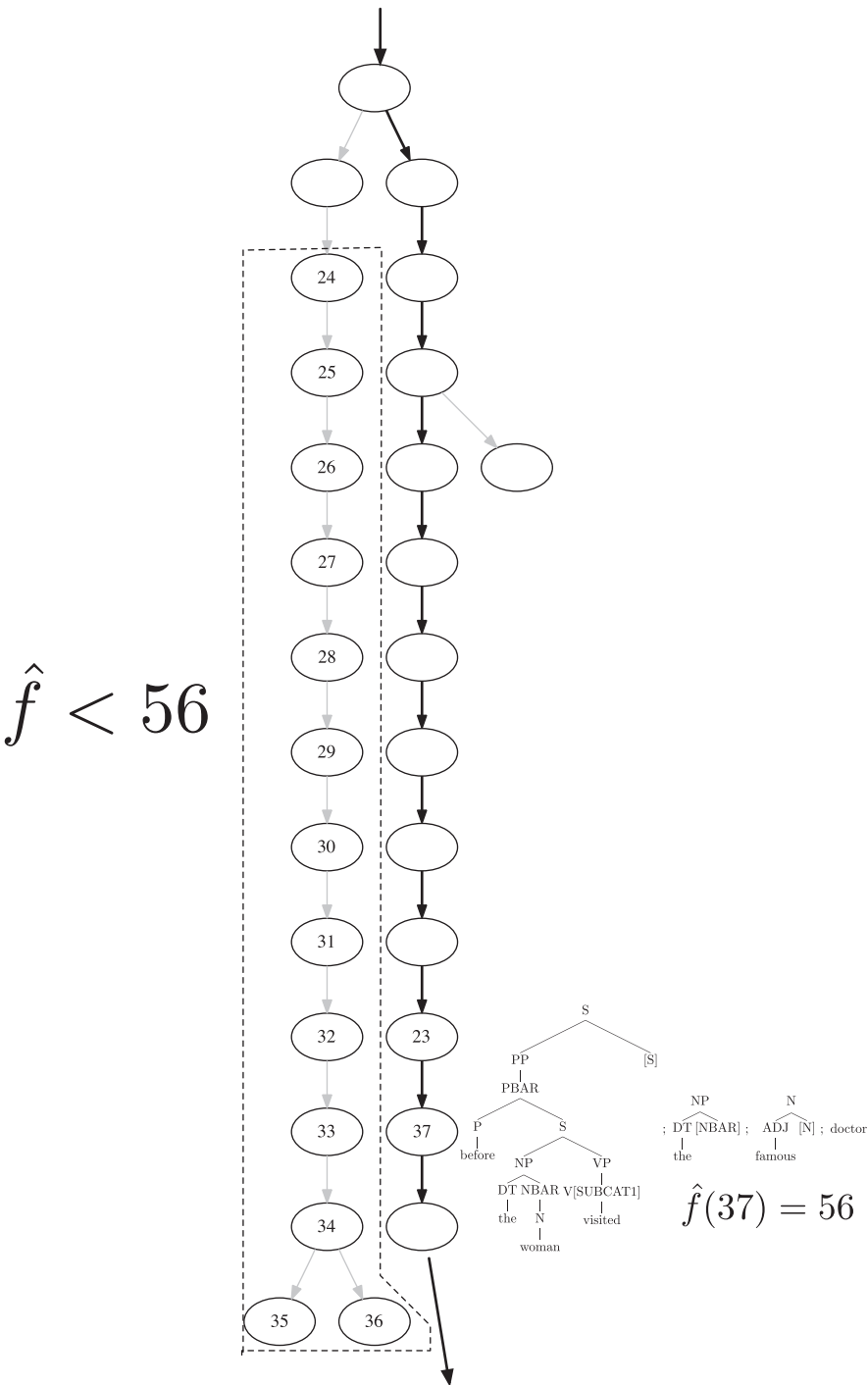


Fig. 24. Problem subspace for the “hard” NP/Z garden-path sentence.

By contrast, the purposive analysis proposed in this article does not assume any sort of cost for settling on a syntactic analysis or escaping from one. These costs would be needed to work out numerical predictions from Chater et al.'s theory. In the broad framework proposed here, everything depends on just one heuristic that estimates the amount of work left to do. This heuristic may be obtained from a corpus sample that a theorist chooses to interpret as a list of parsing successes, or it can be derived from a weighted grammar. In problem subspaces where more alternatives need to be disconfirmed before proceeding, this leads to more work. This approach represents an overall simplification of Chater et al.'s (1998) proposal.

Section 4.3 is in part concerned with local coherence effects studied by Bicknell et al. (2009). These authors suggest that local coherence effects would be rational under the assumption that "the bottom-up prior is available to the human sentence processor more rapidly than the in-context posterior." Within the broad framework, the availability of bottom-up information is a consequence of an explicit parsing strategy that is partially bottom-up. While both proposals agree that bottom-up parsing is happening in these cases, the broad framework supplies a functional motivation for these actions in terms of a pressure to finish the task.

Models in the proposed broad framework span Tabor and colleagues' (2004) distinction between Self-Consistent Parsing Accounts and Local Coherence Accounts. On the one hand, the framework does not require that states be self-consistent in the way that an oracle would enforce. On the other, states do have a categorical interpretation as either containing or not containing particular pieces of phrase structure. It would be possible to interpret the broad framework's heuristic value as degrees-of-coherence; however, this would ignore the fact that under A* search, the heuristic value incorporates predictive guesses about the future.

Other researchers have not explicitly set out to do rational analysis but have made proposals that overlap with the one made in this article. A comparison can be made to the derivational theory of complexity (Miller, 1962). The derivational theory of complexity (DTC) asserts a positive relationship between processing difficulty and the number of syntactic transformations required to derive the sentence. The broad framework proposes a kind of DTC in that each grammar rule has potentially observable processing cost. Unlike the classic DTC however, this cost is not related to syntactic transformations but instead reflects parser operations that directly build surface structure. For instance, the effects of a transformation such as topicalization might be spread across many different phrase structure rules according to a slash-passing metarule (Crain & Fodor, 1985; Levine & Hukari, 2004). Experimental evidence against the classic DTC does not directly bear on the new model because of this difference between rule types. The natural extension of the DTC to phrase structure would associate the length of the globally correct derivation with perceptual difficulty. The broad framework contrasts with this natural extension because, due to temporary ambiguity, more than the minimum number of search steps will generally need to be taken. Nevertheless, because both approaches relate grammar rules to increments of predicted difficulty, the proposal can be viewed as a kind of revision to the DTC.

As mentioned in the introduction, the broad framework is also related to Augmented Transition Networks (ATNs). Psycholinguistic modeling with ATNs has always employed a uniformly top-down strategy. The examples in Sections 4 and 5 demonstrate a mixed

strategy. On the other hand, the complexity metric discussed in Section 3.3 is basically the same as Kaplan's (1972) "number of transitions made or attempted" metric.

The process modeling in this article represents a conservative extension of Frazier and Fodor's Garden-Path Theory. For Frazier and Fodor (1978), pursuing the wrong analysis leads to extra tree-building actions. In the broad framework proposed here, the same is true. But in the spirit of rational analysis, we decline to impose a memory limit. This could be done using any one of a number of methods (Korf, 1995; Russell, 1992; Zhou & Hansen, 2002) that go beyond beam search (Gibson, 1991; Roark & Johnson, 1999). Indeed, with the right heuristic, a memory bound may be unnecessary (Brants & Crocker, 2000).

7. Conclusion

The rational analysis methodology leads to a broad framework for theories of human sentence comprehension, one that makes clear that the process of finding a syntactic analysis is a kind of search. To specify an algorithmic model within the framework, one takes a position on the human parsing strategy and the valuation of intermediate states. A simple formulation, motivated by functional pressures to understand *quickly* and *accurately*, derives garden pathing, garden-pathing counterexamples and local coherence in English and German. Analyzing this simple model yields a refined model whose relation to a grammar is more direct. This refined model aligns with the idea of entropy-reduction: when more alternatives must be ruled out, comparatively more work is done. The predictions of this refined model, as regards the relative size of two garden-path examples, indeed align correctly with previous work. Because the broad framework is situated at Marr's (1982) middle level, it details what a rational parser would do on a step-by-step basis. This approach provides a look inside a mechanism that seems to be compatible with many of the empirical findings and conceptual constraints proposed in the field of sentence comprehension to this point.

Notes

1. The Introduction categorizes a multifaceted body of work along two specific dimensions. These dimensions—the characterization of informational factors that affect human sentence processing, and the definition of general theories of comprehension difficulty—naturally go hand in hand and all of the cited articles make some contribution to both enterprises. For purposes of this Introduction, they are separately identified in order to motivate the sort of analysis sketched in Fig. 5.
2. The notation for intermediate parser states such as those shown in Fig. 6 uses a comma to separate the current stack of phrase structures from the remaining input list. In certain grammars used in this article, the comma is also a terminal symbols (e.g., Figs 1, 20, and 22(a)). This literal comma only appears inside the stack or inside the remaining input list, whereas the separator comma appears between them.

The use of the literal comma is essential because it has played a key role in several experimental contrasts that the comprehension model captures.

3. Some human readers unaccustomed to garden-path sentences are flatly unable to comprehend these constructions. The simple model has the combinatorial potential to model both experienced garden-path readers who can understand such constructions and novice garden-path readers who cannot; a threshold would need to be defined on the number of search steps the model takes before it “gives up.” Setting this threshold at 100 states in this way would model a human reader who cannot comprehend Bever’s classic example. Section 3 discusses other linking theory issues.
4. In order to use integer arithmetic for all the costs involved, the implementation applies the floor function to the average distance estimates.
5. As mentioned above in Section 2.4, the grammars discussed in this article are not lexicalized. Lexicalized grammars (as applied e.g., by Srinivas, Kim, and Trueswell, 2002) appear to be the most promising method for dealing with lexical effects such as those studied by Ford, Bresnan, and Kaplan (1982); MacDonald, Pearlmutter, and Seidenberg (1994); Mitchell (1987). While the extension to lexicalized grammars presents no formal obstacle, the practical estimation problem associated with them is more complex. As the same informed search techniques and parsing strategies could be applied with these grammars, this article leaves such extension to future work.
6. A critic might argue that the proposed comprehension model is a notational variant of Frazier’s (1979) Garden Path model in the sense that $g(n)$ prioritizes the exploration of phrase structures that have fewer nodes first. By setting the step-costs identically to zero, the role of $g(n)$ can be suppressed. The fact that the same pattern of search behavior still arises confirms that $\hat{h}(n)$ has an important role too. This term of the A* heuristic value equation has no parallel in Garden Path theory. This demonstration rules out the possibility that the proposal is a notation variant of the Garden Path model. A more complete comparison, provided in Section 6, details similarities and differences between the broad framework and other proposals in the literature.
7. To the author’s knowledge, Gibson’s judgment about the lack of garden pathing in Example 2 has not proved sufficiently controversial to merit detailed experimental investigation. However, an outcome showing a garden-path effect in one example but not the other might revive interest in theories of fixed attachment preferences.
8. The common noun “ball” shares many semantic and distributional properties with “frisbee” and appears 27 times in the Brown corpus. Because it consists of texts published in 1961, it is perhaps unsurprising that frisbees are not mentioned more often.
9. This procedure is identical to the one described for English in Section 3.2 on page 12.
10. The stochastic expectation matrix $\mathcal{A}$ is also known as the first moment (Everett & Ulam, 1948, p. 5)
11. Fronted PPs are likely even rarer in speech corpora. This weighting illustrates the ability of the refined model to reflect arbitrary factors that may not be reflected in a particular treebank.

Acknowledgments

The author would like to express his appreciation to the reviewers of this manuscript for their kind help, as well as to audiences at the International Workshop on Parsing Technologies, UCLA, Stanford, MIT, Edinburgh, Washington, Cornell, NYU, Yale, Chicago, and Michigan. Thanks are due as well to Jim Fill, Austin Frank, Greg Kobele, Christina Kim, Rick Lewis, David Lutz, Mats Rooth, Helmut Schmid, Ed Stabler, and Whitney Tabor. The research reported in this article was supported by NSF CAREER award number 0741666.

References

- Abney, S., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3), 233–249.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hilldale, NJ: Lawrence Erlbaum Associates.
- Bach, E. (1976). An extension of classical transformational grammar. In *Problems of linguistic metatheory* (pp. 183–224). Available at: <http://people.umass.edu/ebach/papers/extcl.htm>. Accessed on October 21, 2010.
- Bailey, K. G., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, 49(2), 183–200.
- Berwick, R. C., & Weinberg, A. S. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York, NY: Wiley.
- Bicknell, K., Levy, R., & Demberg, V. (2009). Correcting the incorrect: Local coherence effects modeled with prior belief update. In I. Kwon, H. Pritchett & J. Spence (Eds.), *Proceedings of the 35th annual meeting of the Berkeley linguistics society* (pp. 13–24). Berkeley, CA: Berkeley Linguistics Society.
- Boland, J. E. (1993). The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research*, 22(2), 133–152.
- Boston, M. F., Hale, J. T., Patil, U., Kliegl, R., & 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.
- Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*.
- Brants, T., & Crocker, M. (2000). Probabilistic parsing and psychological plausibility. In *Proceedings of 18th international conference on Computational Linguistics COLING-2000*. Saarbrücken/Luxembourg/Nancy: The Association for Computational Linguistics.
- Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirra, S., König, E., Lezius, W., Rohrer, C., Smith, G., & Uszkoreit, H. (2004). TIGER Linguistic interpretation of a German corpus. *Research on Language and Computation*, 2, 597–619.
- Bresnan, J. (1982). Introduction: Grammars as mental representations of language. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. xvii, lii). Cambridge, MA: MIT Press.
- Bresnan, J. (2001). *Lexical-functional syntax*. Malden, MA: Blackwell.
- Carroll, P. J., & Slowiczek, M. L. (1991). Modes and modules: Multiple pathways to the language processor. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 221–247). Cambridge, MA: MIT Press.
- Chambers, C. G., Magnuson, J. S., & Tanenhaus, M. K. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(3), 687–696.
- Chater, N., Crocker, M. W., & Pickering, M. J. (1998). The rational analysis of inquiry: The case of parsing. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 441–468). New York: Oxford University Press.

- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). Some notes on economy of derivation and representation. In *The Minimalist Program* (chap. 2) (pp. 129–166). Cambridge, MA: MIT Press.
- Clifton, C., Speer, S., & Abney, S. P. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30(2), 251–271.
- Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 94–127). Cambridge, England: Cambridge University Press.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 320–538). New York: Cambridge University Press.
- Crocker, M. W. (2005). Rational models of comprehension: Addressing the performance paradox. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 363–380). Mahwah, NJ: Erlbaum.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193–210.
- Demberg, V., & Keller, F. (2009). A computational model of prediction in human parsing: Unifying locality and surprisal effects. In N. Taatgen, & H. van Rijn (Eds.), *Proceedings of the 29th meeting of the Cognitive Science Society* (pp. 1888–1893).
- Demers, A. J. (1977). Generalized left corner parsing. In *Conference report of the 4th annual association for computing machinery symposium on principles of programming languages* (pp. 170–181). New York: Association for Computing Machinery.
- Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the Association for Computing Machinery*, 13(2), 94–102.
- Eisner, J. (2002). Discovering syntactic deep structure via Bayesian statistics. *Cognitive Science*, 26(3), 255–268.
- Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 195–225). Cambridge, MA: MIT Press.
- Everett, C. J., & Ulam, S. M. (1948). *Multiplicative systems in several variables I* (Tech. Rep. No. LA-683). In A. R. Bednarek & F. Ulam (Eds.), *Analogies between analogies: The mathematical reports of S. M. Ulam and his Los Alamos collaborators*. Available at: <http://ark.cdlib.org/ark:/13030/ft9g50091sl>. Accessed on October 21, 2010.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- 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). Cambridge, MA: MIT Press.
- Frank, S. (2010). Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 81–89). Uppsala, Sweden: Association for Computational Linguistics. Available at: <http://www.aclweb.org/anthology/W10-2010>. Accessed on October 26, 2010.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.
- Frazier, L., & Clifton, C. Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–325.
- Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 615–628.
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University.
- 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.

- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research*, 32(2), 141–166.
- Grune, D., & Jacobs, C. J. H. (2008). *Parsing techniques: A practical guide*. (2nd ed.) New York: Springer.
- 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* Available at: <http://www.aclweb.org/anthology/W10-201>. Accessed on October 26, 2010.
- 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.
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions of systems science and cybernetics*, ssc-4(2), 100–107.
- Hemforth, B., Konieczny, L., & Strube, G. (1993). Incremental syntax processing and parsing strategies. In *Proceedings of the 15th annual conference of the Cognitive Science Society* (pp. 539–544). Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, England: Cambridge University Press.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 2/4, 647–670.
- Kaplan, R. M. (1972). Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 3, 77–100.
- Kimball, J. P. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15–48.
- Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual meeting of the Cognitive Science Society* (pp. 1178–1183). Mahwah, NJ: Erlbaum.
- Konieczny, L., & Müller, D. (2006). *Local coherence interpretation in spoken language: Evidence from a visual-world-experiment*. (Architectures and Mechanisms of Language Processing)
- Konieczny, L., & Müller, D. (2007). *Local syntactic coherence interpretation in written and spoken language*. (Talk delivered at the 20th annual CUNY conference on human sentence processing)
- Korf, R. E. (1995). Space-efficient search algorithms. *ACM Computing Surveys*, 27(3), 337–339.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Laird, J. E., & Newell, A. (1983). A universal weak method. In P. S. Rosenbloom, J. E. Laird, & A. Newell (Eds.), *The Soar papers: Research on integrated intelligence (chap. 6)* (pp. 245–292), Cambridge, MA: MIT Press.
- Levine, R., & Hukari, T. E. (2004). *The unity of unbounded dependency constructions (Vol. no. 160)*. Stanford, CA: Center for the Study of Language and Information.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Lewis, R. L. (1993). *An architecturally-based theory of human sentence comprehension*. Unpublished doctoral dissertation, Pittsburgh, PA: Carnegie Mellon University.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.
- Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman and Company.

- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12(2/3), 137–176.
- Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, 17(10), 748–762.
- Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–681). Hillsdale, NJ: Erlbaum.
- Mitchell, D. C., Cuertos, F., Corley, M. M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24(6), 469–488.
- Narayanan, S., & Jurafsky, D. (1998). Bayesian models of human sentence processing. In *Proceedings of the 20th annual conference of the Cognitive Science Society*. Madison, WI: University of Wisconsin-Madison.
- Narayanan, S., & Jurafsky, D. (2002). Bayesian model predicts human parse preference and reading time in sentence processing. In S. B. Thomas G. Dietterich & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14: Proceedings of the 2001 conference* (pp. 59–65). Cambridge, MA: MIT Press.
- Newell, A. (1968). On the analysis of human problem-solving protocols. In J.-C. Gardin & B. Jaulin (Eds.), *Calcul et formalisation dans les sciences de l'homme* (pp. 146–185). Paris: Editions du Central National de la Recherche Scientifique.
- Newell, A. (1980). The knowledge level. In P. S. Rosenbloom, J. E. Laird, & A. Newell (Eds.), *The Soar papers: Research on integrated intelligence* (pp. 136–176). Cambridge, MA: MIT Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Port, R. F. (2003). Dynamical systems hypothesis in cognitive science. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 1027–1032). New York: Nature Publishing Group. doi:10.1002/0470018860.s00020.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America*, 90(6), 2956–2970.
- Pritchett, B. (1988). The grammatical basis of language processing. *Language*, 64(3), 539–576.
- Pritchett, B. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Pylyshyn, Z. W. (1999). What's in your mind? In E. LePore & Z. W. Pylyshyn (Eds.), *What is Cognitive Science?* (pp. 1–20). Malden, MA: Blackwell.
- Resnik, P. (1992). Left-corner parsing and psychological plausibility. In *Proceedings of the fourteenth international conference on computational linguistics*. Nantes, France.
- Roark, B., & Johnson, M. (1999). *Broad coverage predictive parsing*. Presented at the 12th Annual CUNY Conference on Human Sentence Processing.
- Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 324–333). Singapore: Association for Computational Linguistics. Available at: <http://www.aclweb.org/anthology/D/D09/D09-1034>. Accessed on October 26, 2010.
- Rosenkrantz, D. J., & Lewis, P. M. (1970). Deterministic left corner parsing. In *IEEE conference record of the 11th annual symposium on switching and automata* (pp. 139–152). New York: IEEE. DOI: 10.1109/SWAT.1970.5.
- Russell, S. J. (1992). Efficient memory-bounded search methods. In B. Neumann (Ed.), *Proceedings of the Tenth European Conference on Artificial Intelligence* (pp. 1–5). Chichester, NY: Wiley.
- Schabes, Y. (1990). *Mathematical and computational aspects of lexicalized grammars*. Unpublished doctoral dissertation, University of Pennsylvania.
- Schuler, W., Miller, T., AbdelRahman, S., & Schwartz, L. (2010). Broad-coverage parsing using human-like memory constraints. *Computational Linguistics*, 36(1), 1–30.
- Shieber, S., & Johnson, M. (1993). Variations on incremental interpretation. *Journal of Psycholinguistic Research*, 22(2), 287–318.

- Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25(2), 249–271.
- Srinivas, B., Kim, A., & Trueswell, J. (2002). *The convergence of lexicalist perspectives in psycholinguistics and computational linguistics*. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing* (pp. 109–135). Amsterdam: John Benjamins.
- Stabler, E. (1991). Avoid the pedestrian's paradox. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 199–237). Dordrecht, The Netherlands: Kluwer.
- Steedman, M. (1989). Grammar, interpretation and processing from the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 463–504). Cambridge, MA: MIT Press.
- Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.
- Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21(2), 165–201.
- Stowe, L. A. (1989). Thematic structures and sentence comprehension. In G. N. Carlson, & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 319–357). Boston: Kluwer.
- Sturt, P., & Crocker, M. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11(5), 449–494.
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40, 136–150.
- Sturt, P., Costa, F., Lombardo, V., & Frascioni, P. (2003). Learning first-pass structural attachment preferences with dynamic grammars and recursive neural networks. *Cognition*, 88(2), 133–169.
- 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.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Thorne, J. P. (1968). A computer model for the perception of syntactic structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 171(1024), 377–386. Available at: <http://www.jstor.org/stable/75835>. Accessed on October 26, 2010.
- Tutunjian, D., & Boland, J. (2008). Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass*, 2(4), 631–646.
- Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research*, 22(3), 339–364.
- Wetherell, C. (1980). Probabilistic languages: A review and some open questions. *Computing Surveys*, 12(4), 361–379.
- Wilson, K., & Carroll, J. B. (1954). Applications of entropy measures to problems of sequential structure. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research* (pp. 103–110). Bloomington, IN: Indiana University Press.
- Yun, J., Whitman, J., & Hale, J. (2010). Subject-object asymmetries in Korean sentence comprehension. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 2152–2157). Austin, TX: Cognitive Science Society.
- Zhou, R., & Hansen, E. A. (2002). Memory-bounded A* graph search. In S. Haller, & G. Simmons (Eds.), *Proceedings of the fifteenth international FLAIRS conference* (pp. 203–209). Menlo Park, CA: AAAI Press.

Appendix A: Generalization in Example 1

The empirical heuristic discussed in Section 2 falls back to shorter and shorter stacks until it finds one that has “enough” observations associated with it. The first column of each table gives the length of this abstracted substack for which an estimate was available in the Brown corpus. While different variants of this heuristic could be proposed with higher or

lower thresholds, speakers' creative use of language makes renders it inevitable that some sort of generalization procedure will be required.

Experimental			Control		
Prefix Length	Stack	Observations	Prefix Length	Stack	Observations
1	while [TOP]	269	1	while [TOP]	269
2	IN [TOP]	2,308	2	IN [TOP]	2,308
3	[S] SBAR [TOP]	1,090	3	[S] SBAR [TOP]	1,090
3	Mary [S] SBAR [TOP]	3	3	Mary [S] SBAR [TOP]	3
4	NNP [S] SBAR [TOP]	114	4	NNP [S] SBAR [TOP]	114
4	NP [S] SBAR [TOP]	1,151	4	NP [S] SBAR [TOP]	1,151
3	[VP] SBAR [TOP]	1,379	4	[VP] SBAR [TOP]	1,379
4	was [VP] SBAR [TOP]	117	3	[VP] SBAR [TOP]	1,379
4	VBD [VP] SBAR [TOP]	595	4	was [VP] SBAR [TOP]	117
3	[VP] SBAR [TOP]	1,379	4	VBD [VP] SBAR [TOP]	595
3	[PP] SBAR [TOP]	199	3	[VP] SBAR [TOP]	1,379
2	mending [VP] SBAR [TOP]	1	3	[PP] SBAR [TOP]	199
4	VBG [VP] SBAR [TOP]	59	2	mending [VP] SBAR [TOP]	1
3	[NP] SBAR [TOP]	440	4	VBG [VP] SBAR [TOP]	59
2	SBAR [TOP]	1,067	3	[NP] SBAR [TOP]	440
3	[S] S [TOP]	1,025	2	SBAR [TOP]	1,067
4	[.] [S] S [TOP]	73	3	[S] S [TOP]	1,025
4	a [NP] SBAR [TOP]	57	4	[.] [S] S [TOP]	73
4	DT [NP] SBAR [TOP]	210	3	[S] S [TOP]	1,025
3	[NN] SBAR [TOP]	124	4	a [S] S [TOP]	12
2	sock [NN] SBAR [TOP]	1	4	DT [S] S [TOP]	101
2	SBAR [TOP]	1,067	5	[NN] NP [S] S [TOP]	97
3	[S] S [TOP]	1,025	5	sock [NN] NP [S] S [TOP]	1
4	[.] [S] S [TOP]	73	4	NP [S] S [TOP]	1,034
4	a [S] S [TOP]	12	3	[VP] S [TOP]	6,260
4	DT [S] S [TOP]	101	4	fell [VP] S [TOP]	6
5	[NN] NP [S] S [TOP]	97	4	VBD [VP] S [TOP]	2,438
5	sock [NN] NP [S] S [TOP]	1	3	[VP] S [TOP]	6,260
4	NP [S] S [TOP]	1,034	3	[PP] S [TOP]	1,296
3	[VP] S [TOP]	6,260	4	on [PP] S [TOP]	78
4	fell [VP] S [TOP]	6	4	IN [PP] S [TOP]	1,128
4	VBD [VP] S [TOP]	2,438	3	[NP] S [TOP]	2,448
3	[VP] S [TOP]	6,260	4	the [NP] S [TOP]	581
3	[PP] S [TOP]	1,296	4	DT [NP] S [TOP]	1,199
4	on [PP] S [TOP]	78	3	[NN] S [TOP]	731
4	IN [PP] S [TOP]	1,128	4	floor [NN] S [TOP]	2
3	[NP] S [TOP]	2,448	2	S [TOP]	25,292
4	the [NP] S [TOP]	581			
4	DT [NP] S [TOP]	1,199			
3	[NN] S [TOP]	731			
4	floor [NN] S [TOP]	2			
2	S [TOP]	25,292			