

# 1 How hard is parsing?

The grammars Chomsky originally proposed used transformations, but they also involved a “base” of phrase-structure rules. Lets examine just the base for a moment.

S → NP VP  
 NP → the spy  
 NP → the cop  
 NP → NP PP  
 PP → with binoculars  
 VP → saw NP  
 VP → saw NP PP

(Rayner, Carlson, and Frazier, 1983)

Figure 1: PP attachment grammar

- modifiers are sisters
- noun phrase “adjunction” rule says this PP modifies this NP
- verb phrase likewise has optional prepositional phrase (PP) complement

Syntactic analyses suggest semantic content:

- (1) [S [NP the spy] [VP saw [NP the cop] [PP with binoculars]]]  
 binoculars help the spy do his seeing
- (2) [S [NP the spy] [VP saw [NP [NP the cop] [PP with binoculars]]]]  
 binoculars are a distinguishing possession of the cop

Try to parse a real sentence with a phrase-structure grammar

- look how long this took compared to human sentence understanding!
- the order of rule-picking defines a *parsing strategy*
- for realistic grammars, ambiguity is rampant and chronic

Consider the case of a PP modifying an NP:

- (3) a. Put the block [in the box on the table]  
 b. Put the [block in the box] on the table
- (4) a. Put the block in [[the box on the table] in the kitchen]  
 b. Put the block in [the box [on the table in the kitchen ]]  
 c. Put [[the block in the box] on the table] in the kitchen  
 d. Put [the block [in the box on the table]] in the kitchen  
 e. Put [the block in the box] [on the table in the kitchen]

As the number of PPs increases, the number of possible analyses increases as the Catalan numbers

$$\begin{aligned} \text{Cat}_n &= 2, 5, 14, 42, 132, 469, 1430, 4862, \dots \\ &= \binom{2n}{n} - \binom{2n}{n-1} \end{aligned}$$

(Church and Patil, 1982)

## 2 The Garden-Path Model of Human Sentence Processing

### 2.1 Background

Historical setting: the failure of the Derivation Theory of Complexity

DTC review: transformations unwound to give deep structure. Each unwinding takes constant time. However,

! passive vs. active difficulty dependent on prior linguistic context (Greene, 1970)

! ...and on semantic properties of the NPs (Slobin, 1966)

Q. if it's not transformations, what is it?

A. perceptual strategies (Bever, 1970) (Kimball, 1973)

### 2.2 Perceptual Strategies

A perceptual strategy is a heuristic for going directly from the surface string to a semantic representation without using the grammar. These strategies fail on what Bever termed *garden-path* sentences. For instance,

**Strategy D** Any Noun-Verb-Noun (NVN) sequence within a potential internal unit in the surface structure corresponds to “actor-action-object”.

fails on the garden-path sentence

(5) ### the horse raced past the barn ...fell

	actor	=	horse
because it predicts	action	=	raced
	object	=	barn

when in fact “raced” is not the action (“fell” is the action associated with “the horse”). Many of these strategies could be implemented in a kind of parsing machine called an augmented transition network. (Kaplan, 1972) (Wanner and Maratsos, 1978).

### 2.3 The Garden Path Model

Lyn Frazier and Janet Fodor (1978) claimed to unify a range of perceptual strategies by appealing to the phrase structure tree associated with a *single* current analysis.

“an initial syntactic analysis is assigned on the basis of purely structural information”

**Minimal Attachment** Do not postulate any potentially unnecessary nodes<sup>1</sup>

**Late Closure aka Right Association** If grammatically permissible, attach new items into the clause or phrase currently being processed

Consequences of Minimal Attachment

- The main verb reading (figure 2) requires fewer S nodes (1 rather than 2) than does the reduced relative. The feeling of confusion at “fell” is because the parser must go back and re-analyze at that point.
- The PP-attachment ambiguity (from grammar 1) is explained by that fact that the VP attachment (i.e. VP → saw NP PP) uses one less layer of NP recursion, and is therefore preferred.

Consequences of Late Closure

- The direct object reading (vs. intransitive “jogs”) is correctly predicted for

(6) Since Jay always jogs a mile this seems like a short distance to him  
versus

(7) Since Jay always jogs a mile seems like a short distance to him

---

<sup>1</sup>If you know LR theory, MA solve all reduce-reduce conflicts in favor of the reduce that pops the most symbols (Pereira, 1985). LC/RA solve all shift-reduce conflicts in favor of shifting

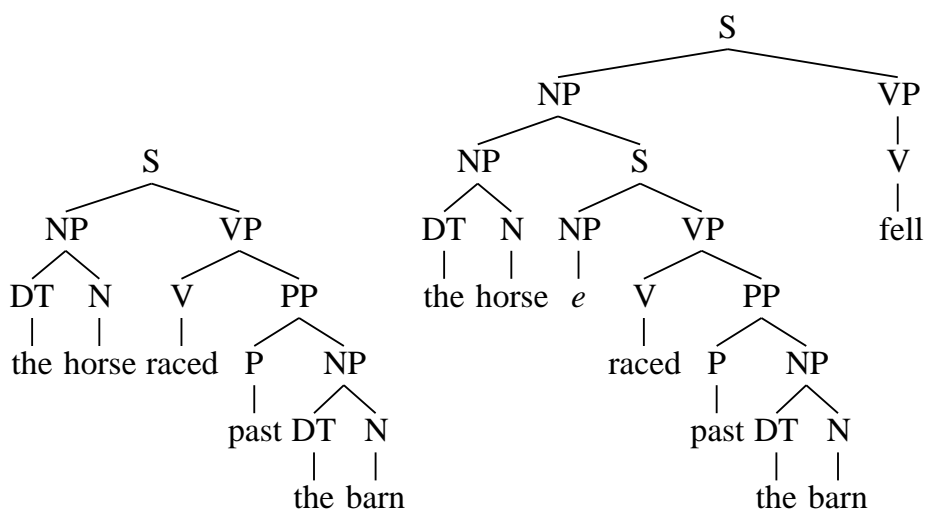


Figure 2: Minimal Attachment prefers the main verb reading

- Low attachment is correctly predicted for adverbs like “yesterday”

- (8) Joyce said Tom left yesterday  
versus  
(9) Joyce said  $E = MC^2$  yesterday

Frazier’s original claim – ‘purely on the basis of structural information’ — is a claim about the modularity of the HSPM. This assertion has been largely disconfirmed by a boatload of evidence that other factors such as referential pragmatics

lexical frequency and combinatory idiosyncrasy

(Crain and Steedman, 1985)  
(Ford, Bresnan, and Kaplan, 1982)  
(MacDonald, Pearlmutter, and Seidenberg, 1994)

all have a measurable effect on ambiguity resolution. that this information constrains processing (perhaps exclusively) is known as *constraint-based lexicalism*. Constraint-based lexicalism emphasizes the fine-grained combinatory and semantic properties of individual words, and has affinities with highly nonmodular *connectionist networks*.

## 2.4 Lexical bias

- (10) Joe bought the book for Susan  
MA predicts that “for Susan” is a beneficiary of “bought”, not a modifier of “the book”
- (11) Joe included my gift for Susan  
Same syntactic structure, but the preferred reading is that “for Susan” is now a modifier of “my gift”

When a set of alternative syntactic categories is reached in the expansion of a PS rule, give priority to the alternatives that are coherent with the strongest lexical form of the verb. (Ford, Bresnan, and Kaplan, 1982)

Animacy (Ferreira and Clifton, 1986) (Trueswell, Tanenhaus, and Garnsey, 1994)

- (12) The witness examined by the lawyer was useless  
(13) The evidence examined by the lawyer was useless

Result: implausibility of evidence examining a lawyer makes garden path effect go away in 13  
Frequency (MacDonald, 1994)

- (14) The rancher knew that the nervous cattle  $\left\{ \begin{array}{l} \text{pushed} \\ \text{moved} \\ \text{driven} \end{array} \right\}$  into the crowded pen were afraid of the cowboys

Result: “pushed” frequency-biased toward transitive leads more ineluctably up the garden path than does “moved” frequency-biased toward intransitive.

Conclusion: there is a potentially huge amount of combinatory information about the “preferences” of individual lexical items (especially verbs) that people act as if they know.

## 2.5 Referential theory

Comprehenders work not with an analysis (“proof”) but with a idea (“model”) about how that world would have to be to make the sentence felicitous.

### Principle of Parsimony

If there is a reading that carries fewer unsatisfied but consistent presuppositions or entailments than any other, then, other criteria of plausibility being equal, that reading will be adopted as most plausible by the hearer, and the presuppositions in question will be incorporated in his or her model. (Crain and Steedman, 1985)

Continuing with the example,

- (15) the horse raced past the barn fell
1. a set of horses  $EQ$  is already in the listener’s discourse model
  2. it’s implicit that the relative clause applies to some individual member of  $EQ$
  3. the whole expression identifies a single individual

an indefinite like

- (16) a horse raced past a barn fell  
presupposes none of these, principle correctly predicts garden-pathing should be reduced.

However, not all processing difficulty would appear to be due to backtracking failed grammatical analyses. Another class involves sentences that are unambiguous.

## 3 Nesting complexity

Perhaps the defining feature of phrase structure grammars is *recursion*. Recursion on the symbol  $S$  is visible in figure 2 where  $S$  is embedded between NP and VP.

Such ‘center embedding’ is harder to comprehend than is right branching or left branching structure (Miller and Chomsky, 1963).

- (17) the dog that chased the cat that saw the rat barked  
versus  
(18) # the cheese that the rat that the cat saw ate stank

Center embedding difficulty is different from garden path difficulty because, in the former but not the later people have no trouble if they are told the intended interpretation. garden path effects are “cognitive penetrable” whereas overload effects like center embedding are not.

It could be that center-embedding is hard because comprehenders have to remember which noun phrase is the argument of which verb; there are at least *function-argument* dependencies between pairs. The first member of the pair must be remembered until the second member is encountered.

**dependency**  $S$  has an  $(i, j)$  dependency with respect to  $L$  if replacement of the  $i^{\text{th}}$  symbol  $a_i$  of  $S$  by  $b_i$  ( $b_i \neq a_i$ ) requires a corresponding replacement of the  $j^{\text{th}}$  symbol of  $S$  by  $b_j$  ( $b_j \neq a_j$ ) for the resulting string to belong to  $L$ . (Chomsky, 1956)

Now, keeping in mind

- the rats eat the cheese  
\* the rat eat the cheese

consider the following three sentences

- (19) the cheese that the rats that the cat saw eat stinks  
(20) \* the cheese that the rat that the cat saw eat stinks (rats  $\rightarrow$  rat)  
(21) the cheese that the rat that the cat saw eats stinks (eat  $\rightarrow$  eats)  
there is an *agreement dependency* between the 5th word rat/rats and the 10th word eat/eats

This leads to a variety of incomplete dependency theories of nesting complexity in which dependencies are described in terms of some formal object in the grammar, such as

phrase structure rules [originally a production theory]	(Yngve, 1960)
theta-grid slots	(Gibson, 1991)
case-assignments	(Stabler, 1994)
or instantiated argument positions	(Johnson, 1998) (Morrill, 2000)

**The Property of Thematic Reception** Associate a load of  $x_{TR}$  PLUs to each C-node constituent that is in a position that can receive a thematic role in some co-existing structure, but whose  $\theta$ -assigner is not unambiguously identifiable in the structure in question. (Gibson, 1991, 97)

Prediction: since [the cheese [that the rat [that the cat ... ] ... ] ...] has three lexical noun phrases that have not been assigned thematic roles, it is unprocessable.

**Bounded connectivity hypothesis** There is a natural typology of linguistic relations such that the psychological complexity of a structure increases quickly when more than one relation of any given type connects a (partial) constituent  $\alpha$  (or any element of  $\alpha$ ) to any constituent external to  $\alpha$ . (Stabler, 1994, 315)

Prediction: since [the cheese [that the rat [that the cat ... ] ... ] ...] has three of the *same kind* of unfulfilled case-assignment relations, it is unprocessable.

Such theories are cannot account for the lack of difficulty in structurally similar items like

- (22) A book [that some Italian [that I never heard of] wrote ] will be published soon by MIT Press (Frank, 1992)
- (23) The reporter who everyone that I met trusts said the president won't resign (Bever, 1974)

in these items, the number of dependencies is the same, but pronouns like “I” are the targets. These are somehow easier. This observation motivates the dependency locality theory.

## 4 The Dependency Locality Theory

**integration cost** The cost of integrating a new head  $h$  is related to the number of new discourse entities introduced between  $h$  and all heads  $h'$  that are in a dependency relation with  $h$ .

**storage cost** (a kind of incomplete dependency metric)

Storage cost at a prefix is related to the minimum number of words necessary to finish off the sentence.

Since pronouns refer to people already in the domain of discourse, they don't count as new discourse entities. In the kind of Davidsonian event-based semantics Gibson implicitly assumes, verbs are also discourse referents.

This predicts the well known subject-object asymmetry by appeal to dependency widths. Center-embedding follows in the same way.

- (24) The reporter<sub>1</sub> who the photographer sent<sub>2</sub>  $e$  to the editor hoped for a good story
- (25) The reporter who<sub>0</sub>  $e$  sent the photographer to the editor hoped for a good story

Storage costs can be manipulated by embedding (Grodner, Gibson, and Tunstall, 2002).

- (26) The witness who the evidence examined by the lawyer implicated was lying

at “examined”, MV stores 1 head (matrix predicate) RR stores 2 or 3 heads (matrix predicate, embedded predicate, gap site for “who”). Plausibility favors RR, but storage favors MV.

Embedding in as a sentence complement serves as a control:

- (27) The witness thought that the evidence examined by the lawyer implicated his next-door neighbor.

Here, storage costs are equal for RR or MV.

Result: ambiguous items read more slowly in the high-storage cost condition.

## References

- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In J.R. Hayes, editor, *Cognition and the Development of Language*. Wiley, New York, pages 279–362.
- Bever, Thomas G., 1974. *Explaining linguistic phenomena*, chapter The ascent of the specious, or there's a lot we don't know about mirrors, pages 173–200. Hemisphere, Washington D.C.
- Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124.
- Church, Kenneth and Ramesh Patil. 1982. Coping with syntactic ambiguity or how to put the block in the box on the table. *Computational Linguistics*, 8:139–149.
- Crain, Stephen and Mark Steedman. 1985. On not being led up the garden path: the use of context by the psychological syntax processor. In David Dowty, Lauri Karttunen, and Arnold Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing. Cambridge University Press, pages 320–538.
- Ferreira, Fernanda and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language*, 25(3):348–368.
- Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan. 1982. A competence-based theory of syntactic closure. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, MA.
- Frank, Robert. 1992. *Syntactic Locality and Tree-adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. Ph.D. thesis, University of Pennsylvania.
- Gibson, Edward. 1991. *A computational theory of human linguistic processing: memory limitations and processing breakdown*. Ph.D. thesis, Carnegie Mellon University.
- Greene, J.M. 1970. The semantic function of negatives and positives. *British Journal of Psychology*, 61:17–22.
- Grodner, Daniel, Edward Gibson, and Susanne Tunstall. 2002. Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, 46:267–295.
- Johnson, Mark. 1998. Proof nets and the complexity of processing center-embedded constructions. *Journal of Logic, Language and Information*, 7(4):433–447.
- Kaplan, Ronald M. 1972. Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 3:77–100.
- Kimball, John P. 1973. Seven principles of surface structure parsing in natural language. *Cognition*, 2:15–48.
- MacDonald, Maryellen C. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, pages 157–201.
- MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg. 1994. Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703.
- Miller, George A. and Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush, and Eugene Galanter, editors, *Handbook of mathematical psychology*. John Wiley, New York, chapter 13, pages 419–491.
- Morrill, Glyn. 2000. Incremental processing and acceptability. *Computational Linguistics*, 26(3).
- Pereira, Fernando. 1985. A new characterization of attachment preference. In David Dowty, Lauri Karttunen, and Arnold Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing. Cambridge University Press, pages 307–319.
- Rayner, Keith, Marcia Carlson, and Lyn Frazier. 1983. The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22:358–374.
- Slobin, Dan I. 1966. Grammatical transformations and sentence comprehension in child hood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 5:219–227.
- Stabler, Edward. 1994. The finite connectivity of linguistic structure. In Charles Clifton, Lyn Frazier, and Keith Rayner, editors, *Perspectives on Sentence Processing*. Lawrence Erlbaum, pages 303–336.
- Trueswell, John C., Michael K. Tanenhaus, and Susan M. Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33:285–318.
- Wanner, Eric and Michael Maratsos. 1978. An ATN approach to comprehension. In Morris Halle, Joan Bresnan, and George A. Miller, editors, *Linguistic Theory and Psychological Reality*. MIT Press, Cambridge, Massachusetts, chapter 3, pages 119–161.
- Yngve, Victor H. 1960. A model and an hypothesis for language structure. In *Proceedings of the American Philosophical Society*, volume 104, pages 444–466.