# Language Science & Technology: Cognitive Foundations II

Matthew W. Crocker

*Department of Computational Linguistics*
*Saarland University*

## Human Language Processing

How do we represent linguistic knowledge

- How are representations stored during comprehension

We understand language incrementally, word-by-word

- How do people construct interpretations

We must resolve local and global ambiguity

- How do people decide upon a particular interpretation

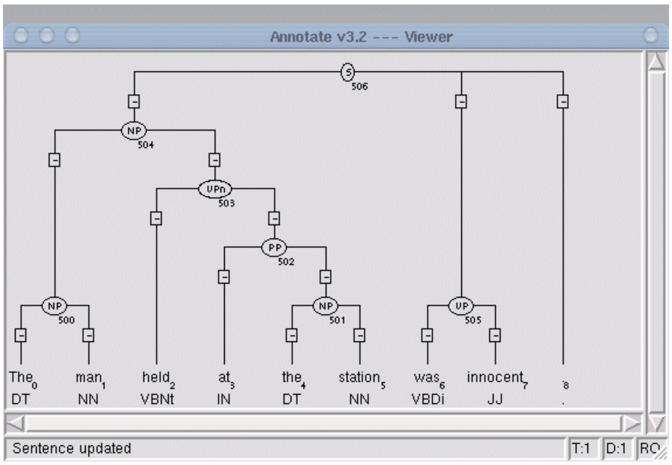
Decisions are sometimes wrong!

- What information is used to identify we made a mistake
- How do we search for an alternative

# The Problem

How do people recover the meaning of an utterance, with respect to a given situation, in real-time?

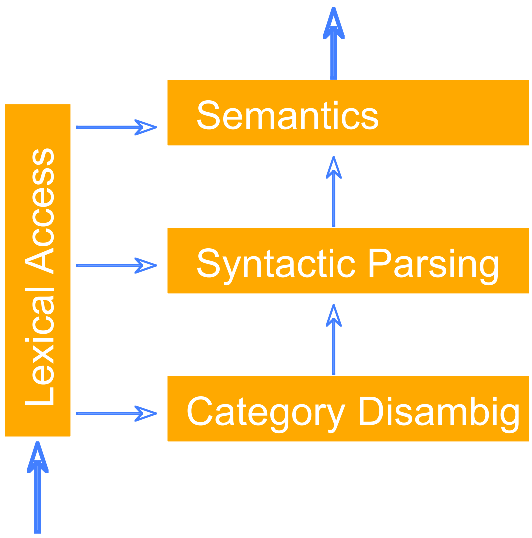
"The man held at the station was innocent"

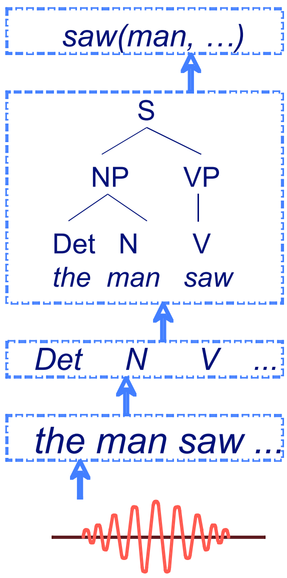



Crocker & Brants, *Journal of Psycholinguistic Research*, 2000.



# A Modular Architecture



# Kind of Mechanisms

Assume we believe that syntactic structure building is underlies sentence comprehension

Questions:

- What kinds of information are used:
  - lexical, grammatical, frequency, semantics, ...
- What kinds of representations:
  - trees, dependencies, AVMs, distributed representations
- What kind of mechanisms:
  - serial/parallel, symbolic/probabilistic/connectionist



# Theories of Sentence Processing

Relate the theory/model to some observed measure

- Typically impossible to predict measures completely

Theories of parsing typically determine …

- what **mechanism** is used to construct interpretations?
- which **information** sources are used by the mechanism?
- which **representation** is preferred/constructed when ambiguity arises?

Linking Hypothesis:

- Preferred sentence structures should have faster reading times in the disambiguating region than dispreferred

## Multiple constraints

"The man/lecture held/fought/given at the station ...

... *a copy of the NY times that he had bought at the airport*" [Main Clause]

... *was rather boring*" [Relative Clause]

Prosody: intonation can assist disambiguation, does it in this case?

Lexical preference: *held* = {Past, **PastPart**}, *fought* = {**Past**, PastPart}, *given* = {PastPart}

Subcat: *held* = { [ _ NP] [ _ NP PP]}, *fought* = { [ _ ] [ _ NP]}
*given* = { [ _ NP PP] [ _ NP NP]}

Semantics: Referential context, plausibility

- Reference: is there more than one man in the context? Yes: prefer relative clause. Why?
- Plausibility: of *man* versus *lecture* as Agent/Patient of the verb



## Two Theories of Sentence Processing

What mechanisms is used to construct interpretations:

- Frazier: Serial parsing, with reanalysis
- Jurafsky: Parallel parsing, with reranking
- McRae: Competitive activation of alternatives

What information is used to determine preferred structure:

- Frazier: General syntactic principles
- Jurafsky: Relative probabilities of alternative structures
- McRae: Competitive integration of constraints

# Overview of the Garden Path Theory

Parsing preferences are guided by general principles:

- Serial structure building
- Reanalyze based on syntactic conflict
- Reanalyze based on low plausibility ("thematic fit")

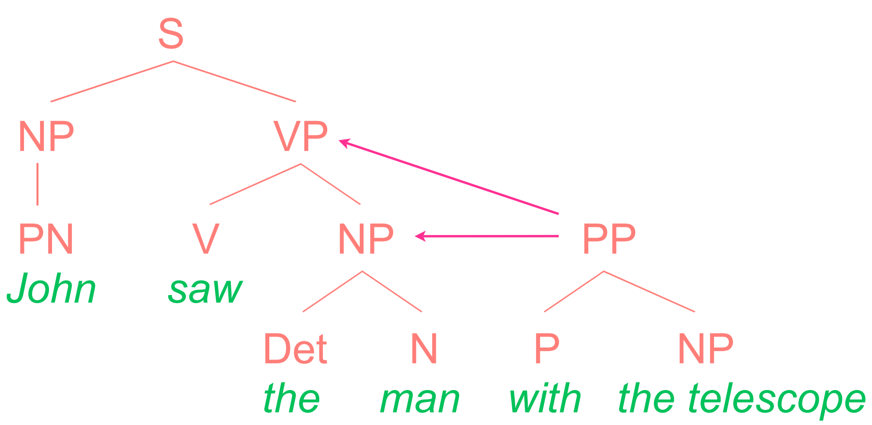
Psychological assumptions:

- Modularity: only syntactic (not lexical, not semantic) information used for initial structure building
- Resources: emphasizes importance of memory limitations
- Processing strategies are universal, innate



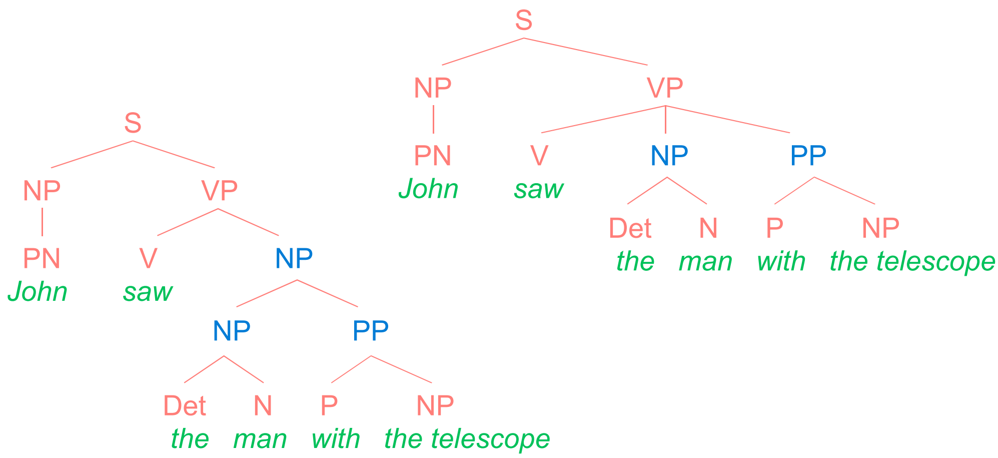
# The Garden Path Theory (Frazier)

Prepositional Phase Attachment:



## First Strategy: Minimal Attachment

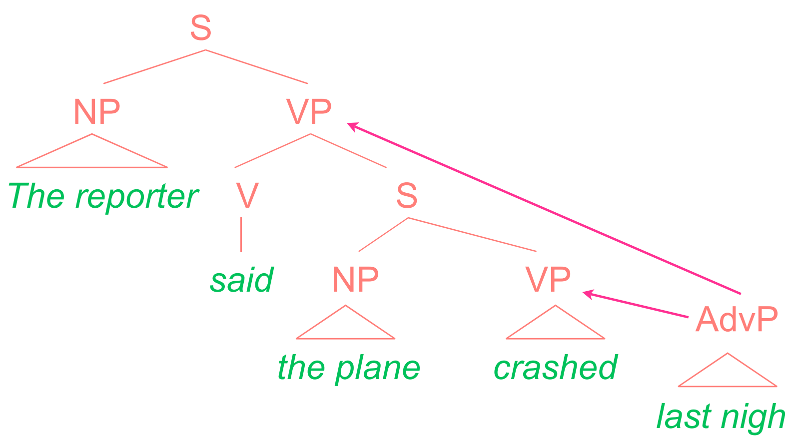
**Minimal Attachment:** Adopt the analysis which requires postulating the fewest nodes






## Second Strategy: Late Closure

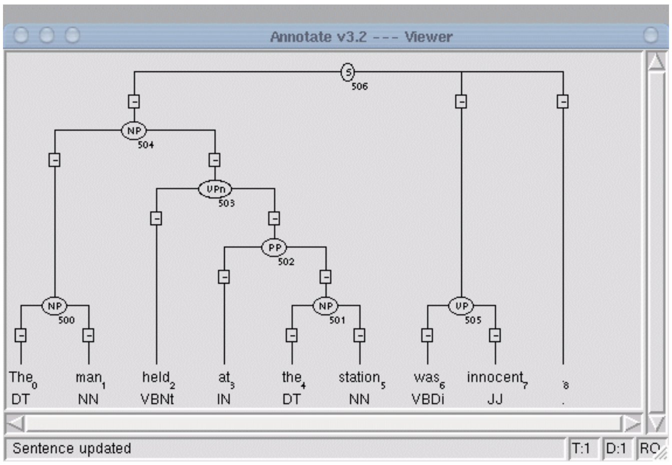
**Late Closure:** Attach material into the most recently constructed phrase marker



# The Problem

How do people recover the meaning of an utterance, with respect to a given situation, in real-time?

"The man held at the station was innocent"




Crocker & Brants, *Journal of Psycholinguistic Research*, 2000.



# Garden-Path Theory: Frazier (1978)

What **architecture** is assumed?

- Modular syntactic processor, with restricted lexical (category) and semantic knowledge

What **mechanisms** is used to construct interpretations?

- Incremental, serial parsing, with reanalysis

What **information** is used to determine preferred structure?

- General syntactic principles based on the current phrase stucture

**Linking Hypothesis**:

- Parse complexity and reanalysis cause increased RTs

## Probabilistic Theories: The Role of Experience

Task of comprehension: recover the correct interpretation

- Goal: Determine the most likely analysis for a given input:

$$\arg\max_i P(s_i) \text{ for all } s_i \in S$$

P can hide a multitude of sins:

- P corresponds to the degree of belief in an interpretation
- Influenced by recent utterances, experience, context

Implementation:

- P is determined by frequencies in corpora or completions
- To compare probabilities (of the Si), assume parallelism



## Implementation

Interpretation of probabilities

- Likelihood of structure occurring, P can be determined by frequencies in corpora or human completions

Estimation of probabilities

- Infinite structural possibilities = sparse data
- Associate probabilities with grammar (finite): e.g. PCFGs

What mechanisms are required:

- Incremental structure building and estimation of probabilities
- Comparison of probabilities entails parallelism

# Probabilistic Grammars

Context-free rules annotated with probabilities

- Probabilities of all rules with the same LHS sum to one;
- Probability of a parse is the product of the probabilities of all rules applied in the parse.

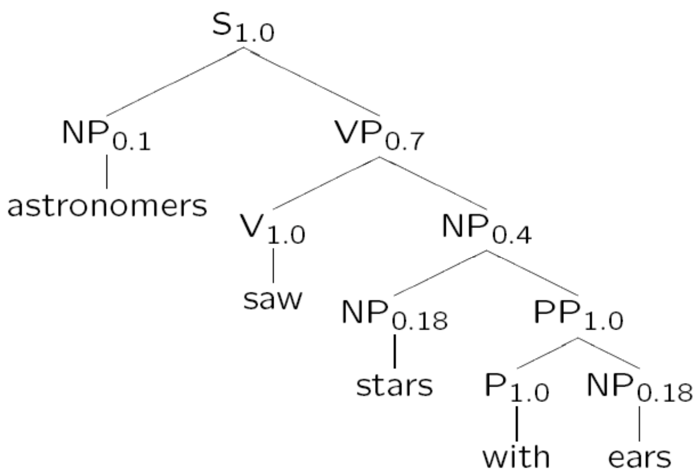
Example (Manning and Schütze 1999):

| | | | |
|---|---|---|---|
| S → NP VP | 1.0 | NP → NP PP | 0.4 |
| PP → P NP | 1.0 | NP → astronomers | 0.1 |
| VP → VP NP | 0.7 | NP → ears | 0.18 |
| VP → VP NP | 0.3 | NP → saw | 0.04 |
| P → with | 1.0 | NP → stars | 0.18 |
| V → saw | 1.0 | NP → telescopes | 0.1 |



# Parse Ranking

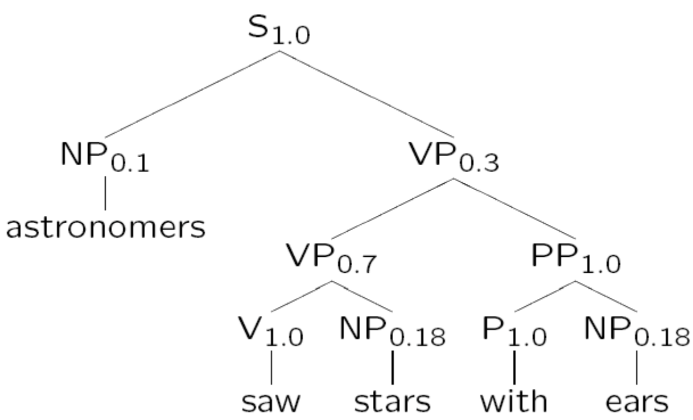
$t_1$:




$$P(t_1) = 1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.0009072$$

# Parse Ranking

$t_2$:




$$P(t_1) = 1.0 \times 0.1 \times 0.3 \times 0.7 \times 1.0 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.0006804$$



# Jurafsky (1996)

Probabilistic model of lexical and syntactic disambiguation

- exploits concepts from computational linguistics:
  - PCFGs, Bayesian modeling frame probabilities.

Overview of issues:

- data to be modeled: frame preferences, garden paths;
- architecture: serial, parallel, limited parallel;
- probabilistic CFGs, frame probabilities;
- examples for frame preferences, garden paths

# Modeling Garden Paths

The reduced relative clause often cause irrecoverable difficulty, but not always:

- The horse raced past the barn fell (irrecoverable)
- The bird found died (recoverable)

Probabilities can distinguish these two cases, in a way a purely structural account (Frazier) cannot.

Assume a *bounded*, parallel parser …

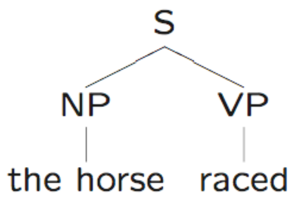
- Only those parsers which are within some "beam" of the preferred parse are kept, others are discarded



# The horse raced past the barn fell

$p(\text{race}, \langle \text{NP} \rangle) = 0.92$

$t_1$:

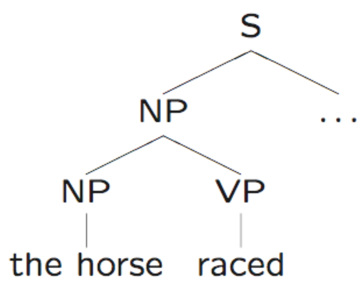



$p(t_1) = 0.92$ (preferred)

$p(\text{race}, \langle \text{NP NP} \rangle) = 0.08$

NP → NP XP 0.14

$t_2$:

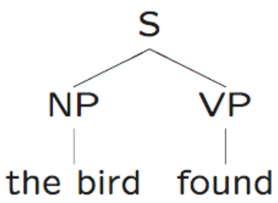



$p(t_1) = 0.0112$ (dispreferred)

# The bird found died

$p(\text{find}, \langle \text{NP} \rangle) = 0.38$

$t_1$:

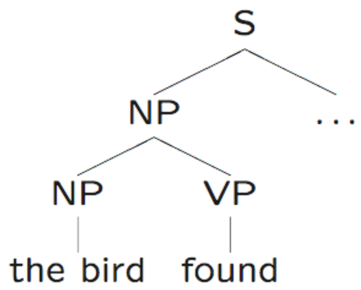



$p(t_1) = 0.38$ (preferred)

$p(\text{find}, \langle \text{NP NP} \rangle) = 0.62$

NP → NP XP   0.14

$t_2$:




$p(t_1) = 0.0868$ (dispreferred)



# The Jurafsky Model

Setting the beam width:

- "The horse raced past the barn fell"  82:1
- "The bird found died"  4:1

Jurafsky assumes a a parse is "pruned" if its probability ratio with the best parse is greater than 5:1

- predicts a garden path for parses that have been pruned

Open issues:

- Where do we get the probabilities?
- Still purely syntactic: what about other constraints?

# Garden-Path Theory: Jurafsky (1996)

What **architecture** is assumed?

- Modular lexico-syntactic processor with lexical (category and subcategory), no semantic knowledge

What **mechanisms** is used to construct interpretations?

- Incremental, bounded parallel parsing, with reranking

What **information** is used to determine preferred structure?

- Lexical and structural probabilities

**Linking Hypothesis**:

- Parse reranking causes increased RTs, if correct parse has been eliminated, predict a garden-path



# The Competitive-Integration Model (McRae et al, 1998)

Claim: Diverse constraints (linguistic and conceptual) are brought to bear simultaneously in ambiguity resolution.

The Model: *Assumes the all analyses are constructed*

- Constraints provide "probabilistic" support for analyses
  - Constraint are weighted and normalized
  - Lexical & structural bias, parafoveal cues, thematic fit ...

Goal: Simulate reading times

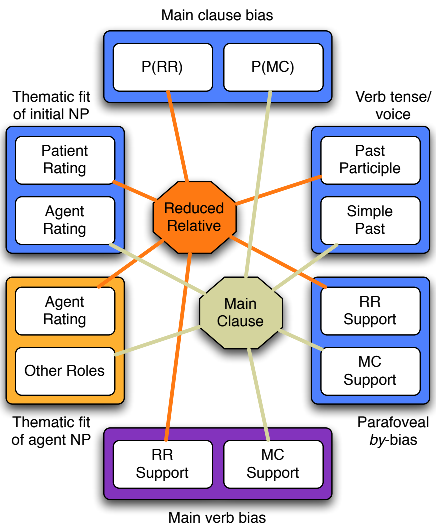
- RTs are claimed to correlate with the number of cycles required to settle on one of the alternatives

"No model-independent signature data pattern can provide definitive evidence concerning when information is used"

# The Computational Model

The crook arrested by the detective was guilty of taking bribes

1. Combines constraints as they become available in the input
2. Input determines the probabilistic activation of each constraint
3. Constraints are weighted according to their strength
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times






# Constraint Parameters

*"The crook/cop arrested by the detective was guilty of taking bribes"*

Verb tense/voice constraint: verb bias towards past or past participle

Relative log frequency is estimated from corpora: RR=.67 MC=.33

Main clause bias: general bias for structure for "NP verb+ed …"

Corpus: P(RR|NP + verb-ed) = .08, P(MC|NP + verb-ed) = .92

*by*-Constraint: extent to which 'by' supports the passive construction

Estimated for the 40 verbs from WSJ/Brown: RR= .8 MC= .2

Thematic fit: the plausibility of crook/cop as an agent or patient

Estimated using a rating study

*by*-Agent thematic fit: good Agent is further support for the RR vs. MC

Same method as (4).

## The recurrence mechanism

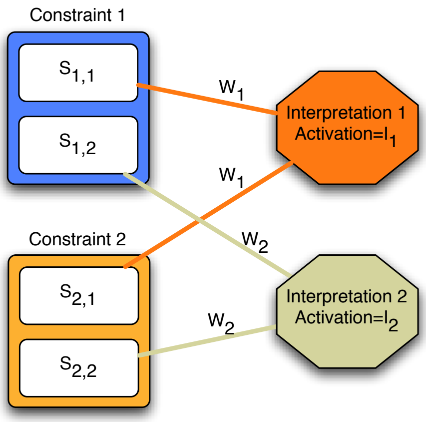
$S_{c,a}$ is the <u>raw activation</u> of the node for the $c^{th}$ constraint, supporting the $a^{th}$ interpretation,

$w_c$ is the <u>weight</u> of the $c^{th}$ constraint

$I_a$ is the activation of the $a^{th}$ interpretation

3-step normalized recurrence mechanism:

- Normalize: $S_{c,a}(norm) = \dfrac{S_{c,a}}{\sum_a S_{c,a}}$
- Integrate: $I_a = \sum_c \left[ w_c \cdot S_{c,a}(norm) \right]$
- Feedback: $S_{c,a} = S_{c,a}(norm) + I_a \cdot w_c \cdot S_{c,a}(norm)$




$$\sum_i w_i = 1$$



## Constraint-based Models

What **architecture** is assumed?

- Non-modular: all levels are constructed and interact simultaneously

What **mechanisms** is used to construct interpretations?

- Parallel: ranking based on constraint activations

What **information** is used to determine preferred structure?

- All relevant information and constraints use immediately

**Linking Hypothesis**:

- Comprehension is easy when constraints support a common interpretation, difficult when they compete

## Summary

People are extremely good at understanding language

- fast, accurate, robust and adaptive to context

There are some "pathologies", where processing is imperfect

- centre-embedding, ambiguity resolution, garden paths
- experimental methods can provide detailed insights

These findings are used to shape the development of models

- serial, parallel, competitive activation
- modular, interactive
- rule-based, knowledge-based or probabilistic

Models make predictions, so we run more experiments!



## For the exam ...

Be familiar with the lecture material

Supplement it with the following two readings:

- Gerry T. M. Altmann. Ambiguity in Sentence Processing. *Trends in Cognitive Sciences*, Vol. 2, Num. 4, 1988.
- Edward Gibson and Neal Perlmutter. Constraints on Sentence Comprehension. *Trends in Cognitive Sciences*, Vol. 2, Num. 7, 1988.

Materials are available from the course homepage