Seminar Adaptive Prozesse
May 21, 2001

Jurafsky's Probabilistic Model of Syntactic Disambiguation

Frank Keller

Computerlinguistik
Universität des Saarlandes

keller@coli.uni-sb.de

Overview

Jurafsky's (1996) approach:

- probabilistic model of lexical and syntactic access and disambiguation;
- accounts for psycholinguistic data using concepts from computational linguistics: probabilistic CFGs, Bayesian modeling frame probabilities;
- focus here: syntactic disambiguation in human sentence processing.

Overview of the lecture:

- data to be modeled: frame preferences, garden paths;
- architecture: serial, parallel, limited parallel;
- probabilistic CFGs, frame probabilities;
- examples for frame preferences, garden paths;
- comparison with other models; problems and issues.

Frame Preferences

(1) The women discussed the dogs on the beach.

a. The women discussed the dogs which were on the beach. (90%)

b. The women discussed them (the dogs) while on the beach. (10%)

(2) The women kept the dogs on the beach.

a. The women kept the dogs which were on the beach. (5%)

b. The women discussed them (the dogs) while on the beach. (95%)

Garden Paths

Main Clause vs. Reduced Relative Ambiguity

(3) # The horse raced past the barn fell.
(4) # The teachers taught by the Berlitz method passed the test.
(5) ? The children taught by the Berlitz method passed the test.

Lexical Category Ambiguity

(6) # The complex houses married and single students and their families.
(7) # The warehouse fires destroyed all the buildings.
(8) # The warehouse fires a dozen employees each year.
(9) # The prime number few.
(10) # The old man the boats.
(11) # The grappling hooks on to the enemy ship.

Frame Ambiguity

(12) # The landlord painted all the walls with cracks.
(13) # Ross baked the cake in the freezer.

Parser Architectures

Serial Parser

- build parse trees through successive rule selection;
- if more than one rule applies (choice point), chose one possible tree based on a selection rule;
- if the tree turns out to be impossible, return to the choice point (backtracking) and reparse from there;
- example for selection rule: minimal attachment (choose the tree with the least nodes).

Parallel Parser

- build parse trees through successive rule selection;
- if more than one rule applies, create a new tree for each rule;
- pursue all possibilities in parallel;
- if one turns out to be impossible, drop it;
- problem: number of parse trees can grow exponentially.
- solution: bounded parallelism, only pursue a limited number of possibilities (prune trees).

Modeling Human Parsing

Serial Parser

- garden path means: wrong tree was selected at a choice point;
- backtracking occurs, causes increased processing times.

Parallel Parser

- garden path means: correct tree was pruned;
- backtracking occurs, causes increased processing times.

Jurafsky (1996) assumes bounded parallelism in a parsing model based on probabilistic CFGs.

Pruning occurs if a parse tree is sufficiently improbably (beam search algorithm).

Probabilistic Context-free Grammars

- context-free rules annotated with probabilities;
- probabilities of all rules with the same lefthand side sum to one;
- probability of a parse is the product of the probabilities of all rules applied in the parse.

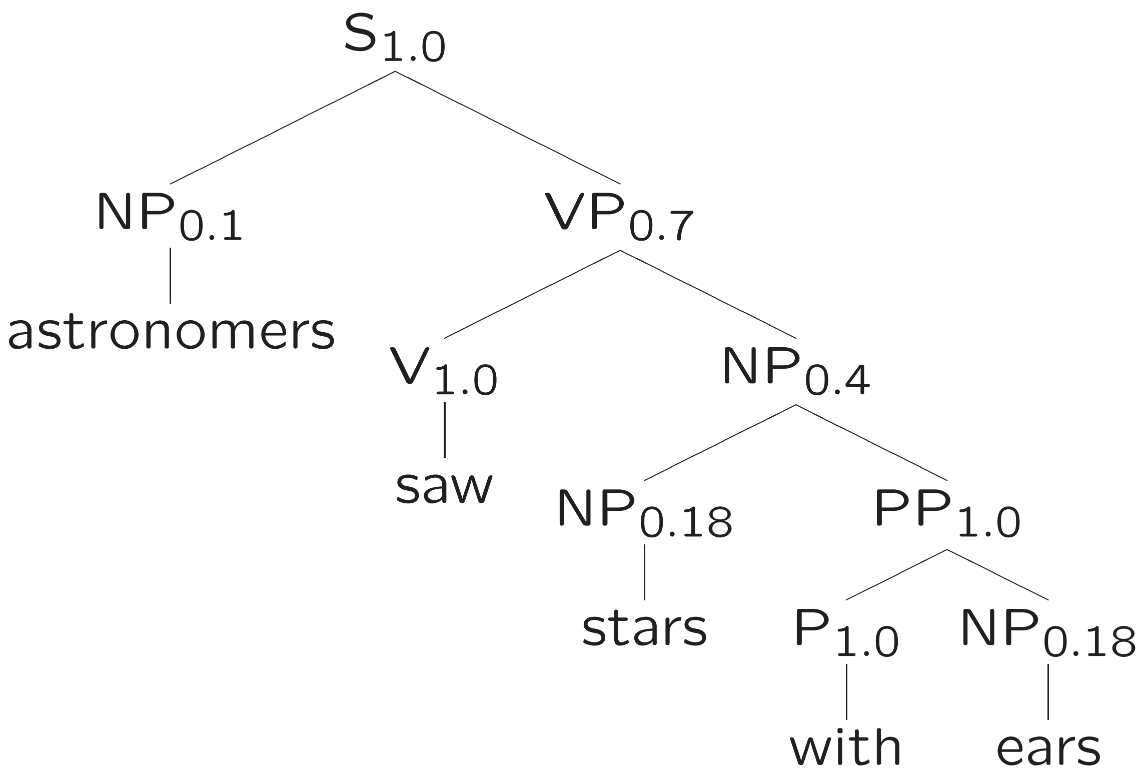
Example (Manning and Schütze 1999)

S $\rightarrow$ NP VP	1.0	NP $\rightarrow$ NP PP	0.4
PP $\rightarrow$ P NP	1.0	NP $\rightarrow$ astronomers	0.1
VP $\rightarrow$ V NP	0.7	NP $\rightarrow$ ears	0.18
VP $\rightarrow$ VP PP	0.3	NP $\rightarrow$ saw	0.04
P $\rightarrow$ with	1.0	NP $\rightarrow$ stars	0.18
V $\rightarrow$ saw	1.0	NP $\rightarrow$ telescopes	0.1

Probabilistic Context-free Grammars

Example (Manning and Schütze 1999)

t_1:

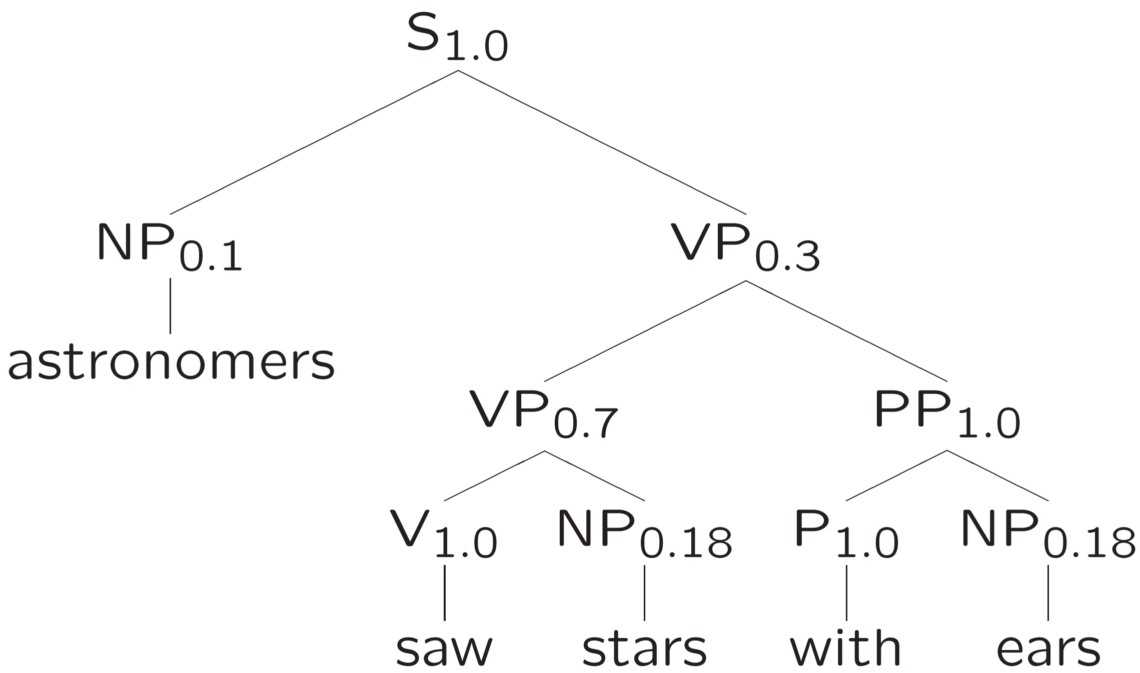


$$P(t_1) = 1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \times 0.18 \times 1.0 \times 1.0 \times 0.18 = 0.0009072$$

Probabilistic Context-free Grammars

Example (Manning and Schütze 1999)

t_2:



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

Frame Probabilities

Complements of *keep*:

AP	keep the prices reasonable
VP	keep his foes guessing
VP	keep their eyes peeled
PRT	keep the people in
PP	keep his nerves from jangling

Frame probabilities computed from the Penn Treebank:

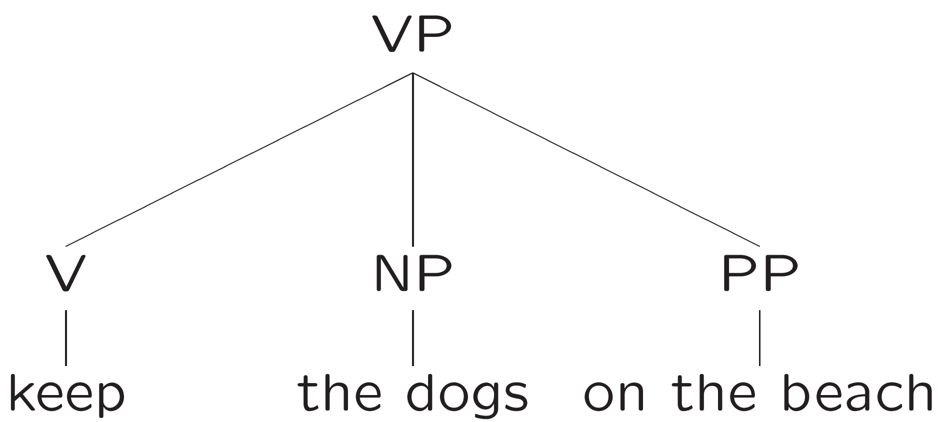
discuss	$\langle$NP PP$\rangle$	.24
	$\langle$NP$\rangle$	.76
keep	$\langle$NP XP[pred +]$\rangle$	.81
	$\langle$NP$\rangle$	.19

Modeling Frame Preferences

$p(\text{keep}, \langle \text{NP XP[pred +]} \rangle) = 0.81$

VP $\rightarrow$ V NP XP   0.15

t_1:



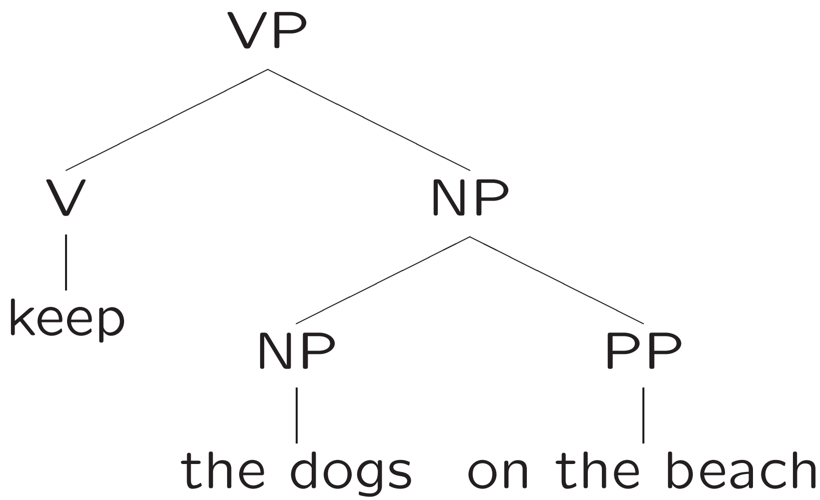
$p(t_1) = 0.15 \times 0.81 = 0.12$ (preferred)

Modeling Frame Preferences

$p(\text{keep}, \langle \text{NP} \rangle) = 0.19$

VP $\rightarrow$ V NP		0.39
NP $\rightarrow$ NP XP		0.14

t_2:



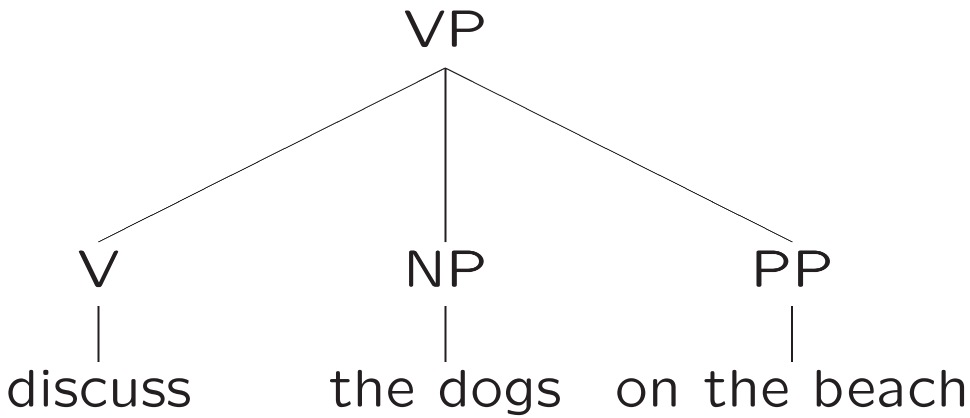
$p(t_2) = 0.19 \times 0.39 \times 0.14 = 0.01$ (dispreferred)

Modeling Frame Preferences

$p(\text{discuss}, \langle \text{NP PP} \rangle) = 0.24$

VP $\rightarrow$ V NP XP   0.15

t_1:



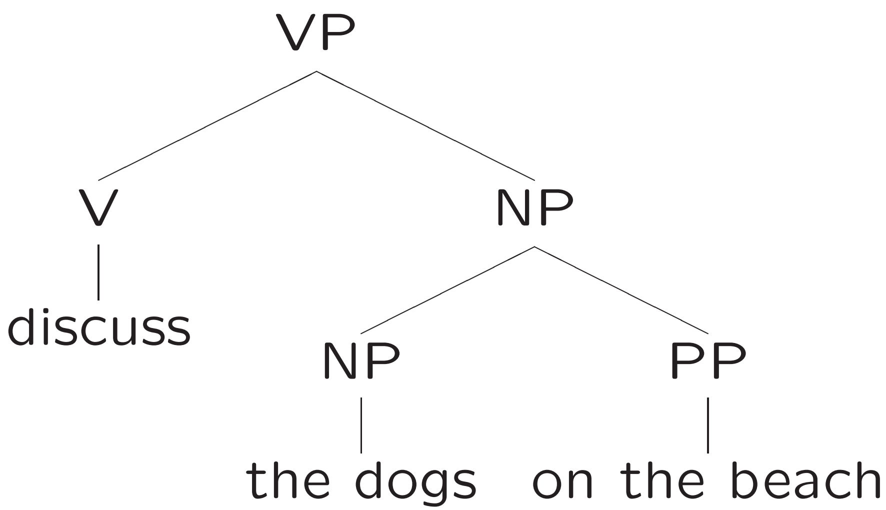
$p(t_1) = 0.15 \times 0.24 = 0.036$ (dispreferred)

Modeling Frame Preferences

$p(\text{discuss}, \langle \text{NP} \rangle) = 0.76$

VP	$\rightarrow$ V NP	0.39
NP	$\rightarrow$ NP XP	0.14

t_2:



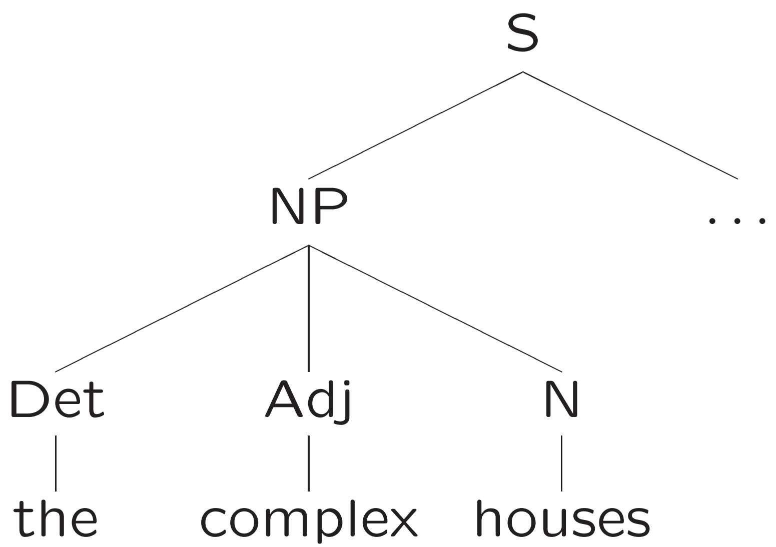
$p(t_2) = 0.76 \times 0.39 \times 0.14 = 0.041$ (preferred)

Modeling Garden Path Effects

Garden path caused by construction probabilities

S $\rightarrow$ NP ...	0.92
NP $\rightarrow$ Det Adj N	0.28
N $\rightarrow$ ROOT s	0.23
N $\rightarrow$ house	0.0024
Adj $\rightarrow$ complex	0.00086

t_1:

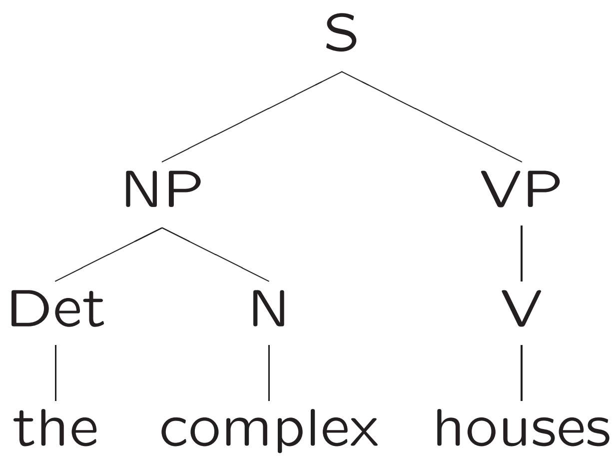


$p(t_1) = 1.2 \times 10^{-7}$ (preferred)

Modeling Garden Path Effects

NP $\rightarrow$ Det N	0.63
S $\rightarrow$ [NP $_{\text{VP}}$[V ...	0.48
N $\rightarrow$ complex	0.000029
V $\rightarrow$ house	0.0006
V $\rightarrow$ ROOT s	0.086

t_1:

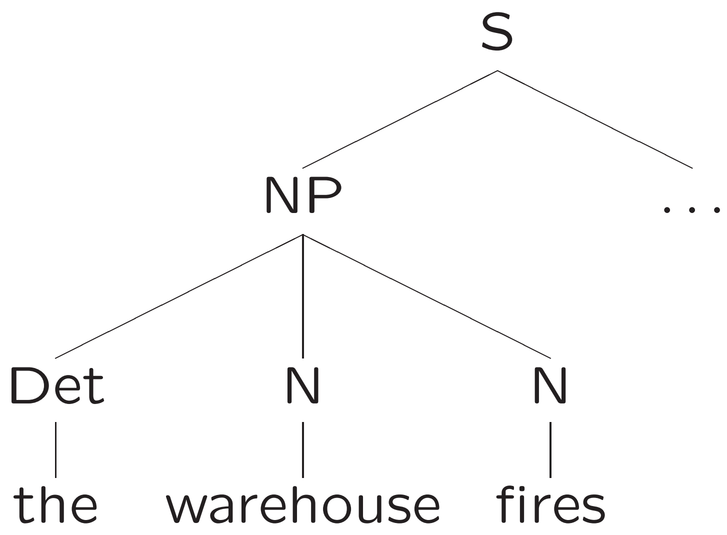


$p(t_1) = 4.5 \times 10^{-10}$ (dispreferred)

Modeling Garden Path Effects

S $\rightarrow$ NP ...	0.92
NP $\rightarrow$ Det N N	0.28
N $\rightarrow$ fire	0.00072
N $\rightarrow$ ROOT s	0.23

t_1:

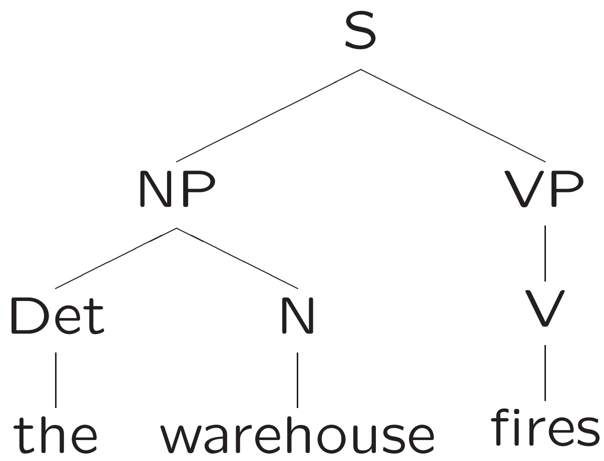


$p(t_1) = 4.2 \times 10^{-5}$ (preferred)

Modeling Garden Path Effects

NP $\rightarrow$ Det N	0.63
S $\rightarrow$ [NP $_{\text{VP}}$[V ...	0.48
V $\rightarrow$ fire	0.00042
V $\rightarrow$ ROOT s	0.086

t_1:



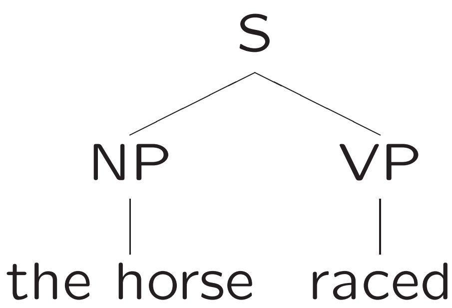
$p(t_1) = 1.1 \times 10^{-5}$ (dispreferred)

Modeling Garden Path Effects

Garden path caused by construction probabilities and frame probabilities

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

t_1:



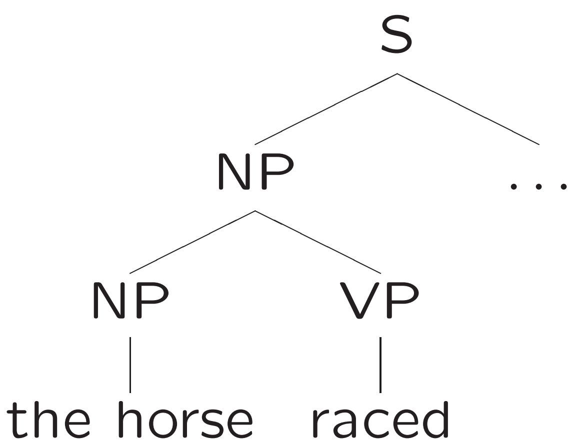
$p(t_1) = 0.92$ (preferred)

Modeling Garden Path Effects

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

NP $\rightarrow$ NP XP   0.14

t_2:

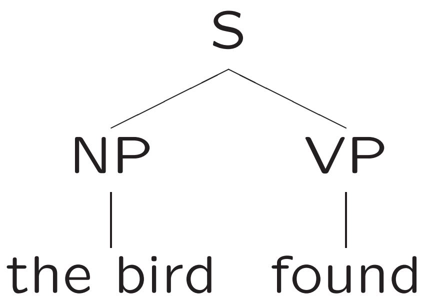


$p(t_1) = 0.0112$ (dispreferred)

Modeling Garden Path Effects

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

t_1:



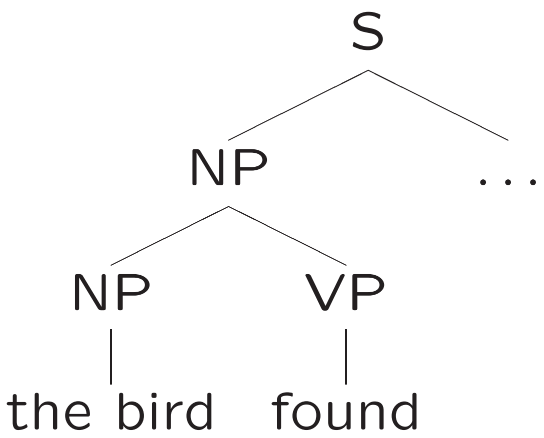
$p(t_1) = 0.38$ (preferred)

Modeling Garden Path Effects

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

NP $\rightarrow$ NP XP 0.14

t_2:



$p(t_1) = 0.0868$ (dispreferred)

Setting the Beam Width

Crucial assumption: if the relative probability of a tree falls below a certain value, then it will be pruned.

sentence	probability ratio
the complex houses ...	267:1
the horse raced ...	82:1
the warehouse fires ...	3.8:1
the bird found ...	3.7:1

Assumption: a garden path occurs if the probability ratio is higher than 5:1.

Open Issues

- Incrementality: Can we make more fine-grained predictions of the time course of ambiguity resolution?
- Coverage: Jurafsky used hand-crafted examples. Can we use a probabilistic parser that is trained on a real corpus?
- Memory Limitations: How can we augment the model to take memory limitations into account (e.g., center embedding)?
- Crosslinguistics: does this model work for languages other than English

References

Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20: 137–194.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.