Probabilistic Models of Human Parsing

Informatics 2A: Lecture 23

Mirella Lapata (slides by Frank Keller)

School of Informatics
University of Edinburgh
mlap@inf.ed.ac.uk

November 10, 2011

Overview

In this lecture, we will discuss a classic probabilistic model of human parsing (Jurafsky, 1996):

- the model integrates lexical and syntactic access and disambiguation;
- it accounts for psycholinguistic data using concepts from NLP: probabilistic CFGs, Bayesian modeling, frame probabilities;
- here, we focus on: syntactic disambiguation in human parsing.

See previous lecture for background on human parsing (garden paths, parser architectures).

Garden Paths

Main Clause vs. Reduced Relative Ambiguity

(1) a. ?The horse raced past the barn fell.
    b. ?The teachers taught by the Berlitz method passed the test.
    c. The children taught by the Berlitz method passed the test.

Frame Ambiguity

(2) a. ?The landlord painted all the walls with cracks.
    b. ?Ross baked the cake in the freezer.

Note: ? means garden path.

Garden Paths

Lexical Category Ambiguity

(3)
a. ?The complex houses married and single students and their families.
b. ?The warehouse fires destroyed all the buildings.
c. ?The warehouse fires a dozen employees each year.
d. ?The prime number few.
e. ?The old man the boats.
f. ?The grappling hooks on to the enemy ship.

Frame Preferences

A verb can have several subcategorization frames (phrases it selects for). Some frames are preferred over others:

(4) 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%)

(5) The women kept the dogs on the beach.
a. The women kept the dogs which were on the beach. (5%)
b. The women kept them (the dogs) while on the beach. (95%)

Results from rating study by Ford et al. (1982).

Clicker Question (1)

Which one of the following is *not* a plausible architecture for a human parser?

1. A serial parser maintains only one analysis at a time
2. A parallel parser maintains several analyses
3. A parser that computes analyses sentence-by-sentence
4. A parser that combines serial processing with limited parallelism

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).

Parser Architectures

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 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 improbable (beam search algorithm).

Probabilistic Context-free Grammars

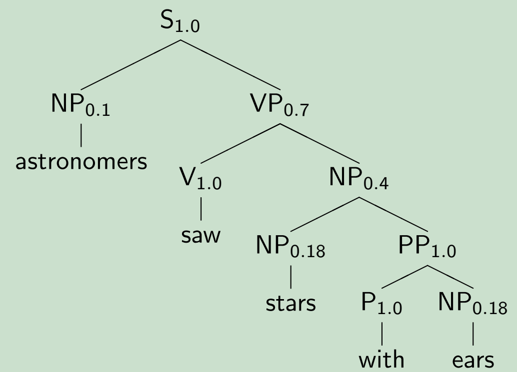
- Context-free rules annotated with probabilities;
- probabilities of all rules with the same lefthand side sum to one;
- probability of a parse is the product of the probabilities of all rules applied in the parse.

Example

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

Probabilistic Context-free Grammars

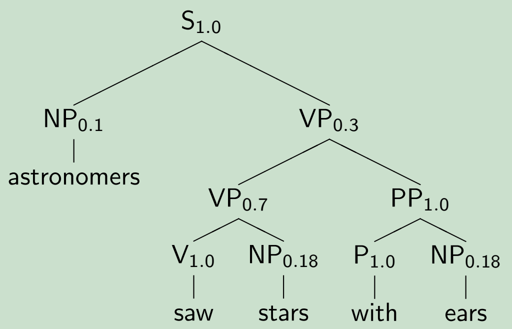
Example



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

Probabilistic Context-free Grammars

Example



$P(t_2) = 1.0 \cdot 0.1 \cdot 0.3 \cdot 0.7 \cdot 1.0 \cdot 0.18 \cdot 1.0 \cdot 1.0 \cdot 0.18 = 0.0006804$

t_1 more probable than t_2: improbable analyses can be pruned.

Frame Probabilities

Subcategorization frames of the verb *keep:*

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

Frame probabilities tell us how likely each of these frames is. This information can be combined with construction probabilities generated by a probabilistic CFG.

Frame Probabilities

Problem: how can frame probabilities be computed?

Solution: use a corpus that's annotated with tree structures (Penn Treebank); estimate frame probabilities from the corpus.

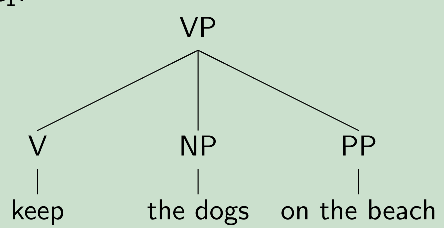
Example

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

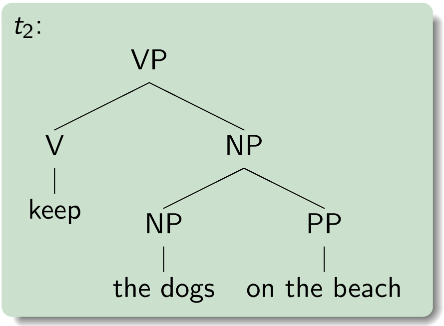
t_1:



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

Modeling Frame Preferences

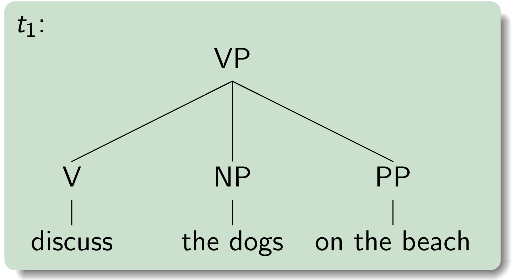
$p(\text{keep}, \langle\text{NP}\rangle) = 0.19$ VP → V NP 0.39
NP → NP XP 0.14



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

Modeling Frame Preferences

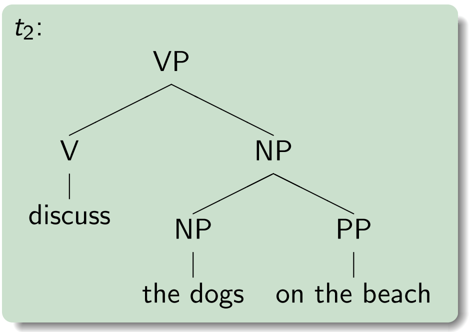
$p(\text{discuss}, \langle\text{NP PP}\rangle) = 0.24$
VP → V NP XP 0.15



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

Modeling Frame Preferences

$p(\text{discuss}, \langle\text{NP}\rangle) = 0.76$ VP → V NP 0.39
NP → NP XP 0.14

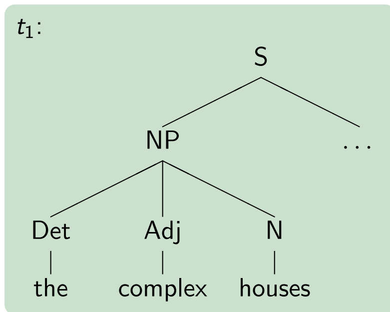


$p(t_2) = 0.76 \cdot 0.39 \cdot 0.14 = 0.041$ (preferred)

Modeling Garden Path Effects

Garden path caused by construction probabilities:

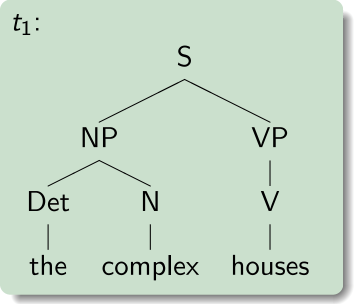
S → NP . . .	0.92	N → house	0.0024
NP → Det Adj N	0.28	Adj → complex	0.00086
N → ROOT s	0.23		



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

Modeling Garden Path Effects

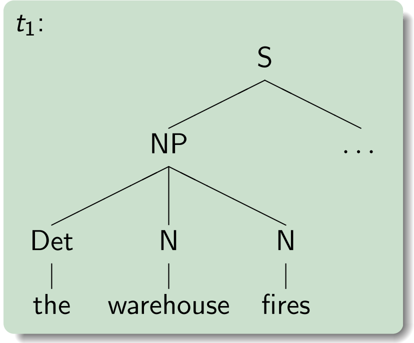
NP → Det N	0.63	V → house	0.0006
S → [NP $_{VP}$[V …	0.48	V → ROOT s	0.086
N → complex	0.000029		



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

Modeling Garden Path Effects

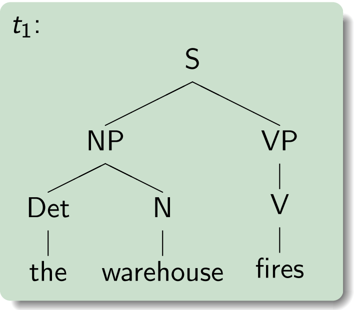
S → NP …	0.92	N → fire	0.00072
NP → Det N N	0.28	N → ROOT s	0.23



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

Modeling Garden Path Effects

NP → Det N	0.63	V → fire	0.00042
S → [NP $_{VP}$[V …	0.48	V → ROOT s	0.086

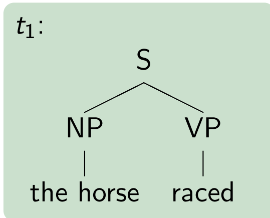


$p(t_1) = 1.1 \cdot 10^{-5}$ (dispreferred)

Modeling Garden Path Effects

Garden path caused by construction probabilities and frame probabilities:

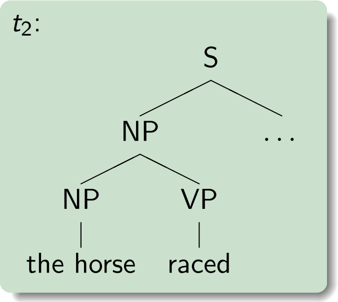
$p(\text{race}, \langle \text{NP} \rangle) = 0.92$



$p(t_1) = 0.92$ (preferred)

Modeling Garden Path Effects

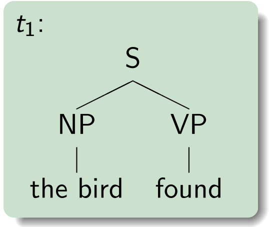
$p(\text{race}, \langle \text{NP NP} \rangle) = 0.08$
NP $\rightarrow$ NP XP   0.14



$p(t_1) = 0.0112$ (dispreferred)

Modeling Garden Path Effects

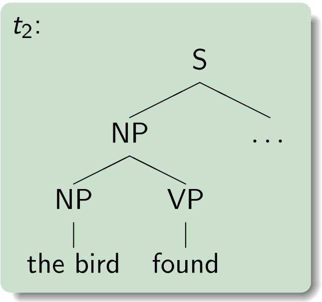
$p(\text{find}, \langle \text{NP} \rangle) = 0.38$



$p(t_1) = 0.38$ (preferred)

Modeling Garden Path Effects

$p(\text{find}, \langle \text{NP NP} \rangle) = 0.62$
NP $\rightarrow$ NP XP   0.14



$p(t_1) = 0.0868$ (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.

Clicker Question (2)

Which one following frames is *least likely* for the verb *drink* ?

1. *The patient must drink several liters each day*
2. *We were up drinking all night*
3. *Let's drink to the New Year*
4. *The mother drinks in every word of her son on the stage*

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)?
- Crosslinguistic validity: does this model work for languages other than English?

Summary

- Different types of garden paths: main clause/reduced relative; frame ambiguity; lexical category;
- rating studies provide evidence for subcat frame preferences;
- modeling assumption:
  - parser with bounded parallelism;
  - pruning of improbable analyses (beam search);
  - probabilistic context-free grammar;
  - subcat frame probabilities;
- Model accounts for different types of garden paths:
  - caused by frame probabilities;
  - caused by construction probabilities;
  - caused by a combination of both;
- beam width: ratio of the probability of the preferred analysis to the dispreferred analysis; needs to be determined empirically.

References