

Introduction to Computational Linguistics

Section

Olga Zamaraeva
University of Washington
May 8, 2020

Plan for today: Preparing for Assignment 5

- ▶ Assignment 4?
- ▶ Projects?
- ▶ PCFG review
- ▶ Making sense of the format
- ▶ Programming concepts
- ▶ Demo

LING472

Section

Plan

PCFG

Format

Unix

Python

PCFG

- ▶ Augment each production with probability that LHS will be expanded as RHS
 - ▶ $P(A \rightarrow B)$ or $P(A \rightarrow B|A)$, $P(\text{RHS}|\text{LHS})$
 - ▶ Sum over all possible expansions is 1
 - ▶ $\sum_{\beta} P(A \rightarrow \beta) = 1$
- ▶ A PCFG is consistent if sum of probabilities of all sentences in language is 1.
 - ▶ Recursive rules can yield inconsistent grammars
 - ▶ We look at consistent grammars in this class

LING472

Section

Plan

PCFG

Format

Unix

Python

Example: PCFG probabilities

(Assume some small CFG grammar of the kind we saw before.)

I have a cat. Do you have a cat? I have a dog.

- ▶ What is $P(\text{Det} \rightarrow a)$?
- ▶ What is $P(\text{N} \rightarrow \text{cat})$?
- ▶ What is $P(\text{S} \rightarrow \text{NP VP})$?

Example: PCFG probabilities

(Assume some small CFG grammar of the kind we saw before.)

$$S \rightarrow \text{aux } S$$

I have a cat. Do you have a cat? I have a dog.

▶ What is $P(\text{Det} \rightarrow a)$?

1 (100%)

▶ What is $P(N \rightarrow \text{cat})$?

$\frac{2}{3}$ (or $\frac{1}{3}$?)

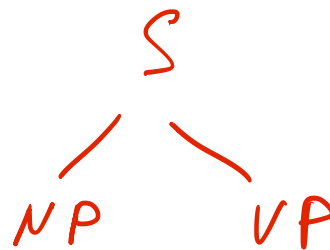
$$\frac{C(N \rightarrow \text{cat})}{C(N)}$$

▶ What is $P(S \rightarrow \text{NP VP})$?

$\frac{2}{3}$

PCKY grammar segment

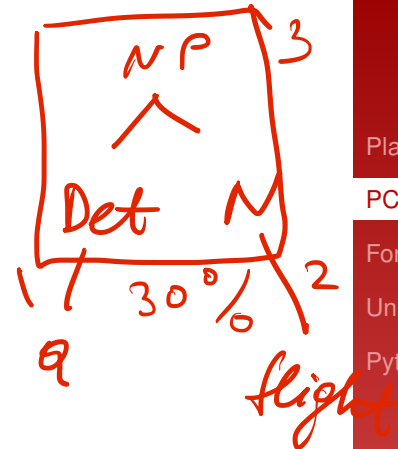
- ▶ $S \rightarrow NP VP$ [0.80]
- ▶ $NP \rightarrow Det N$ [0.30]
- ▶ $VP \rightarrow V NP$ [0.20]
- ▶ $V \rightarrow includes$ [0.05]
- ▶ $Det \rightarrow the$ [0.40]
- ▶ $Det \rightarrow a$ [0.40]
- ▶ $N \rightarrow flight$ [0.02] ←
- ▶ $N \rightarrow meal$ [0.05]



Probabilistic CKY

This flight includes a meal

0	The	1	flight	2	includes	3	a	4	meal	5
Det: .40	NP: .30 * .40 * .02 = .0024									
[0,1]	[0,2]	[0,3]	[0,4]	[0,5]						
	N: .02	X	X	X						
	[1,2]	[1,3]	[1,4]	[1,5]						
		V: .05	X	VP						
		[2,3]	[2,4]	[2,5]						
			Det: .40							
			[3,4]	[3,5]						
				N: .01						
				[4,5]						



0.20×0.05
x

$0.30 \times$
 $0.40 \times$
 0.01

LING472
Section
Plan
PCFG
Format
Unix
Python