LING5702: Lecture Notes 19
A Model of Grammar Acquisition

So far we've seen how babies can discover words in a language.

Today we'll see how (probabilistic) syntactic grammars and lexicons can be learned.

Contents

19.1 A model of grammar acquisition [Jin et al., 2021]

As a baby, you would be exposed to lots of sentences in your caregivers' language.

Imagine you could generate **random grammars** and then assign probabilities to these sentences.

The grammar that best predicts the sentences has the highest probability given the sentences:

$$\begin{aligned}
\mathsf{P}(\text{grammar} \mid \text{sentences}) &= \frac{\mathsf{P}(\text{sentences,grammar})}{\mathsf{P}(\text{sentences})} \\
&= \frac{\mathsf{P}(\text{grammar}) \cdot \frac{\mathsf{P}(\text{sentences,grammar})}{\mathsf{P}(\text{grammar})}}{\mathsf{P}(\text{sentences})} \\
&= \frac{\mathsf{P}(\text{grammar}) \cdot \mathsf{P}(\text{sentences} \mid \text{grammar})}{\mathsf{P}(\text{sentences})} \\
&= \frac{\mathsf{P}(\text{grammar}) \cdot \sum_{\text{trees}} \mathsf{P}(\text{trees} \mid \text{grammar}) \cdot \mathsf{P}(\text{sentences} \mid \text{trees})}{\mathsf{P}(\text{sentences})}
\end{aligned}$$

This is called **Bayes' law**.

Since the probability of grammars is uniform and of sentences given trees is deterministic (based on matching tree terminals), then the tree probabilities determine which grammar is preferred.

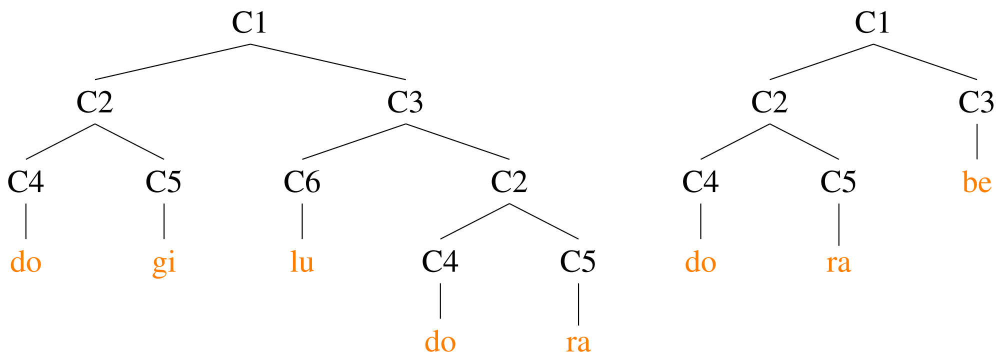
19.2 Example: choosing among grammars

Suppose we encounter the following utterances:

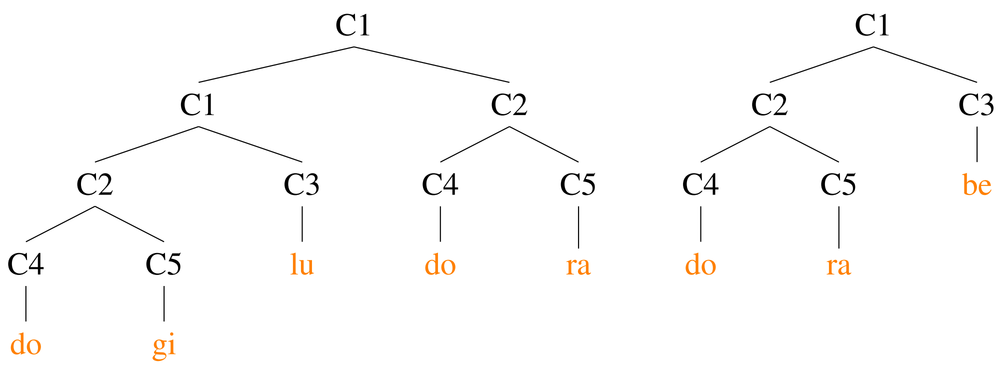
(1) do gi lu do ra

(2) do ra be

We could assign them this analysis (Analysis A):



or we could assign them this analysis (Analysis B):



We can distinguish these based on their probability, according to a probabilistic grammar.

We estimate the rule probabilities using relative frequency estimation:

$$\mathsf{P}(a \rightarrow b\ c \mid a) = \frac{\text{number of times } a \rightarrow b\ c \text{ occurs}}{\text{number of times } a \text{ occurs}}.$$

This gives us the following probabilistic grammar for Analysis A:

$$\mathsf{P}(C1 \rightarrow C2\ C3 \mid C1) = \frac{2}{2} = 1$$

$$\mathsf{P}(C2 \rightarrow C4\ C5 \mid C2) = \frac{3}{3} = 1$$

$$\mathsf{P}(C3 \rightarrow C6\ C2 \mid C3) = \frac{1}{2} = .5$$

$$\mathsf{P}(C3 \rightarrow \text{be} \mid C3) = \frac{1}{2} = .5$$

$$\mathsf{P}(C4 \rightarrow \text{do} \mid C4) = \frac{3}{3} = 1$$

$$\mathsf{P}(C5 \rightarrow \text{gi} \mid C5) = \frac{1}{3} = .333$$

$$\mathsf{P}(C5 \rightarrow \text{ra} \mid C5) = \frac{2}{3} = .667$$

$$\mathsf{P}(C6 \rightarrow \text{lu} \mid C6) = \frac{1}{1} = 1$$

so the total ('joint') probability of all the trees in Analysis A is:

$$\overbrace{1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot .5}^{\text{grammatical rules}} \cdot \overbrace{.5 \cdot 1 \cdot 1 \cdot 1 \cdot .333 \cdot .667 \cdot .667 \cdot 1}^{\text{lexical rules}} = 0.03703$$

(NOTE: probabilities for branches that occur muliple times must be mulitplied in multiple times!)

On the other hand, we get the following probabilistic grammar for Analysis B:

$$\mathsf{P}(C1 \to C2\ C3 \mid C1) = \frac{1}{3} = .333$$
$$\mathsf{P}(C1 \to C1\ C2 \mid C1) = \frac{2}{3} = .667$$
$$\mathsf{P}(C2 \to C4\ C5 \mid C2) = \frac{3}{3} = 1$$
$$\mathsf{P}(C3 \to \text{lu} \mid C3) = \frac{1}{2} = .5$$
$$\mathsf{P}(C3 \to \text{be} \mid C3) = \frac{1}{2} = .5$$
$$\mathsf{P}(C4 \to \text{do} \mid C4) = \frac{3}{3} = 1$$
$$\mathsf{P}(C5 \to \text{gi} \mid C5) = \frac{1}{3} = .333$$
$$\mathsf{P}(C5 \to \text{ra} \mid C5) = \frac{2}{3} = .667$$

so the total ('joint') probability of all the trees in Analysis B is:

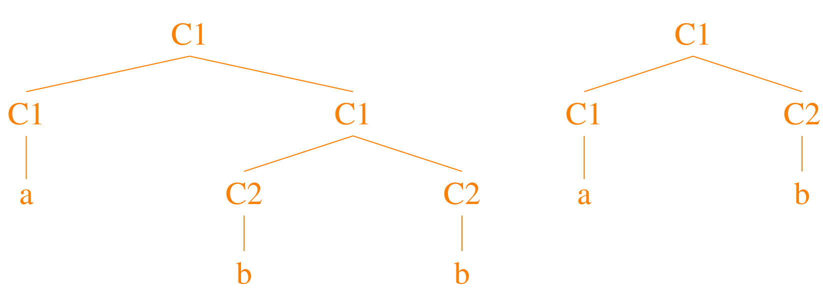
$$\overbrace{.333 \cdot .667 \cdot .667 \cdot 1 \cdot 1 \cdot 1}^{\text{grammatical rules}} \cdot \overbrace{.5 \cdot .5 \cdot 1 \cdot 1 \cdot 1 \cdot .333 \cdot .667 \cdot .667}^{\text{lexical rules}} = 0.005388$$

(NOTE: probabilities for branches that occur muliple times must be mulitplied in multiple times!)

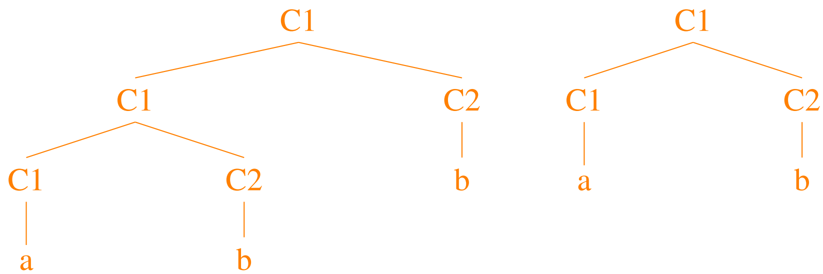
The first analysis is about 7 times more likely!

19.3 Practice

1. Calculate a probabilistic grammar based on the below evidence:



2. Calculate a probabilistic grammar based on the below evidence:



19.4 Practice

Which of the tree sets in the above problem has a lower probability?

References

[Jin et al., 2021] Jin, L., Schwartz, L., Doshi-Velez, F., Miller, T., and Schuler, W. (2021). Depth-Bounded Statistical PCFG Induction as a Model of Human Grammar Acquisition. *Computational Linguistics*, 47(1):181–216.