Exercise 1.
With the grammar

\[
\begin{align*}
  s &\rightarrow np \ vp \\
  vp &\rightarrow v \ np \ pp \\
  vp &\rightarrow v \ np \\
  np &\rightarrow n \\
  np &\rightarrow n \ pp \\
  pp &\rightarrow p \ np
\end{align*}
\]

the sentence "Delis serve pizza with relish." gets two parses. Suppose we were given a training corpus of 5 sentences, with their parses:

\[
\begin{align*}
  (s \ (np \ (n \ Men) \ (pp \ (p \ of) \ (np \ (n \ distinction)))) \ (vp \ (v \ like) \ (np \ (n \ broccoli)))) \\
  (s \ (np \ (n \ Men)) \ (vp \ (v \ like) \ (np \ (n \ ham) \ (pp \ (p \ with) \ (np \ (n \ eggs)))))) \\
  (s \ (np \ (n \ Men)) \ (vp \ (v \ serve) \ (np \ (n \ ham) \ (pp \ (p \ with) \ (np \ (n \ eggs)))))) \\
  (s \ (np \ (n \ Men)) \ (vp \ (v \ serve) \ (np \ (n \ eggs)) \ (pp \ (p \ with) \ (np \ (n \ gusto)))))) \\
  (s \ (np \ (n \ Men)) \ (vp \ (v \ serve) \ (np \ (n \ eggs)) \ (pp \ (p \ to) \ (np \ (n \ customers))))))
\end{align*}
\]

Suppose we used these five parses to train a probabilistic CFG. What probability would be assigned to each production? What probability would be assigned to the two parses for "Delis serve pizza with relish."? In your calculation, consider only the probabilities of the productions; we are not concerned with the probabilities of generating specific lexical items.

Solution

\[
\begin{align*}
  s &\rightarrow np \ vp \ (5 \ out \ of \ 5) \ p = 1.0 \\
  vp &\rightarrow v \ np \ pp \ (2 \ out \ of \ 5) \ p = 0.4 \\
  vp &\rightarrow v \ np \ (3 \ out \ of \ 5) \ p = 0.6 \\
  np &\rightarrow n \ (12 \ out \ of \ 15) \ p = 0.8 \\
  np &\rightarrow n \ pp \ (3 \ out \ of \ 15) \ p = 0.2 \\
  pp &\rightarrow p \ np \ (5 \ out \ of \ 5) \ p = 1.0
\end{align*}
\]
Probability = 1.0 \cdot 0.8 \cdot 0.6 \cdot 0.2 \cdot 1.0 \cdot 0.8 = 0.0768

Probability = 1.0 \cdot 0.8 \cdot 0.4 \cdot 0.8 \cdot 1.0 \cdot 0.8 = 0.2048

The second parse has higher probability, so it will be chosen.

Exercise 2.
The textbook presents a form of probabilistic lexicalized CFG in which both the left and right sides of a production are lexicalized, and then point out that that’s not practical – it produces too many different productions to be able to train a grammar from a treebank. We will consider a more limited use of lexical probabilities in which the probability of applying a production $A \rightarrow BC$ is conditioned on the head of $A$.

(a) Would the conclusion change if the probability of the expansion of the vp node were conditioned on the head of the vp? In other words, if we used the conditional probability $p(r(n)|n, h(n))$ only for the productions expanding vp. Show how you obtained your conclusion (show the revised probabilities for the vp productions and the two parses). [See the end of Lecture 8 / beginning of Lecture 9 lecture notes.]

(b) Make the same comparison (between non-lexicalized and lexicalized probabilities) for the two parses of the sentence "Men like pizza with relish.”
Solution

(a) The revised PCFG vp productions are:

\[ \text{vp(like)} \rightarrow v \ np \ (\frac{2}{3}) \]
\[ \text{vp(like)} \rightarrow v \ np \ pp \ (\frac{0}{2}) \]
\[ \text{vp(serve)} \rightarrow v \ np \ (\frac{1}{3}) \]
\[ \text{vp(serve)} \rightarrow v \ np \ pp \ (\frac{2}{3}) \]

First parse:

\[ \text{Probability} = 1 \cdot 0.8 \cdot \frac{1}{3} \cdot 0.2 \cdot 1 \cdot 0.8 = 0.04266667 \]

Second parse:

\[ \text{Probability} = 1 \cdot 0.8 \cdot \frac{2}{3} \cdot 0.8 \cdot 1 \cdot 0.8 = 0.34133333 \]

The second parse has still higher probability, so it will be chosen again.

(b) "Men like pizza with relish" gets parsed in exact same way as "Delis serve pizza with relish". Two identical parses are produced, so the non-lexicalized probabilities will also be the same, respectively 0.0768 and 0.2048.

\[ \text{(s (np (n Men)) (vp (v like) (np (n pizza) (pp (p with) (np (n relish))))))} \]
\[ \text{Probability} = 1.0 \cdot 0.8 \cdot 0.6 \cdot 0.2 \cdot 1.0 \cdot 0.8 = 0.0768 \]

\[ \text{(s (np (n Men)) (vp (v like) (np (n pizza)) (pp (p with) (np (n relish))))))} \]
\[ \text{Probability} = 1.0 \cdot 0.8 \cdot 0.4 \cdot 0.8 \cdot 1.0 \cdot 0.8 = 0.2048 \]

Regarding lexicalized probabilities we have the following:

First parse:

\[ \text{Probability} = 1.0 \cdot 0.8 \cdot 1 \cdot 0.2 \cdot 1.0 \cdot 0.8 = 0.128 \]

Second Parse:

\[ \text{Probability} = 1.0 \cdot 0.8 \cdot 0 \cdot 0.2 \cdot 1.0 \cdot 0.8 = 0 \]

This time the first parse will be chosen.