COMS W4705: Natural Language Processing (Fall 2018)
Problem Set #2

Wenbo Gao - wg2313@columbia.edu

October 4, 2018

Problem 1 - PCFGs and HHMs

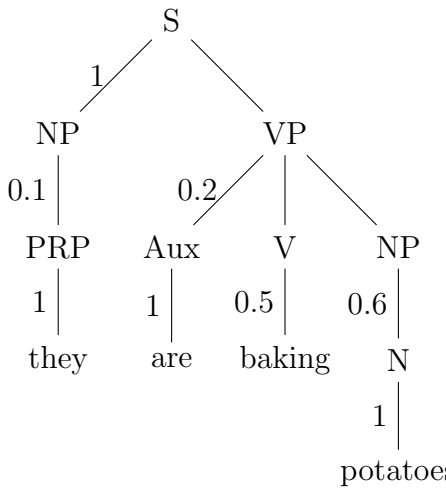
Both PCFGs and HMMs can be seen as generative models that produce a sequence of POS tags and words with some probability (of course the PCFG will generate even more structure, but it will also generate POS tags and words).

(a)

Problem. Revisit the example sentence "*they are baking potatoes*" and grammar from Problem 2. For each sequence of POS tags that is possible for this sentence according to the grammar, what is the joint probability $P(\text{tags}, \text{words})$ according to the PCFG? Hint: consider all parses for the sentence.

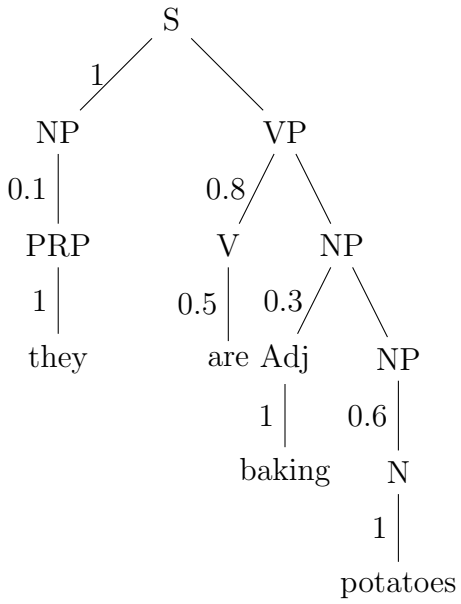
Solution.

first parse:



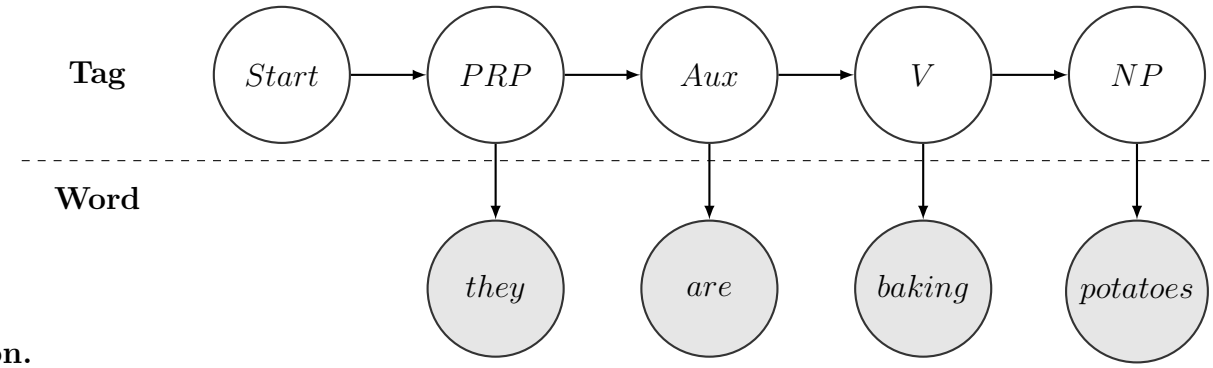
$$
\begin{aligned}
&P[tags, words = \{they, are, baking, potatoes\}] \\
=&P[S \to NP\ VP] \cdot P[NP \to PRP] \cdot P[VP \to Aux\ V\ NP] \cdot P[PRP \to they] \\
&\cdot P[Aux \to are] \cdot P[V \to baking] \cdot P[NP \to N] \cdot P[N \to potatoes] \\
=&1 \cdot 0.1 \cdot 0.2 \cdot 1 \cdot 1 \cdot 0.5 \cdot 0.6 \cdot 1 \\
=&0.006
\end{aligned}
$$

second parse:



$$
\begin{aligned}
&P[tags, words = \{they, are, baking, potatoes\}] \\
=&P[S \to NP\ VP] \cdot P[NP \to PRP] \cdot P[VP \to V\ NP] \cdot P[PRP \to they] \\
&\cdot P[V \to are] \cdot P[NP \to Adj\ NP] \cdot P[Adj \to baking] \cdot P[NP \to N] \cdot P[N \to potatoes] \\
=&1 \cdot 0.1 \cdot 0.8 \cdot 1 \cdot 0.5 \cdot 0.3 \cdot 1 \cdot 0.6 \cdot 1 \\
=&0.0072
\end{aligned}
$$

(b)



Solution.

Problem 2