

# Chapter 14 Statistical Constituency Parsing

9/30/20

5:10 PM

## Probabilistic Parsing

- Compute probability of each interpretation and choose the most probable interpretation

## Probabilistic Context Free Grammar (PCFG)

- Each rule (CFG) is associated with a probability

## PCFG

- Formal Definition

$N$  a set of **non-terminal symbols** (or **variables**)  
 $\Sigma$  a set of **terminal symbols** (disjoint from  $N$ )  
 $R$  a set of **rules** or productions, each of the form  $A \rightarrow \beta$  [ $p$ ],  
where  $A$  is a non-terminal,  
 $\beta$  is a string of symbols from the infinite set of strings  $(\Sigma \cup N)^*$ ,  
and  $p$  is a number between 0 and 1 expressing  $P(\beta|A)$   
 $S$  a designated **start symbol**

- Considering all possible expansions of a non-terminal, sum of all possibilities should be 1.  
If this is true, then the PCFG is **consistent**

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$