

**Al Aho**

**aho@cs.columbia.edu**

# Teaching Compilers



COMPUTER SCIENCE AT  
COLUMBIA UNIVERSITY

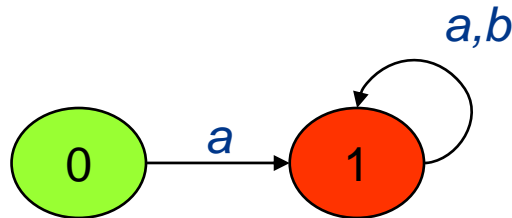
**SIGCSE**  
**Milwaukee, WI**  
**March 12, 2010**

# Specifying Syntax: Regular Expressions and Finite Automata

**Regular expressions** generate the regular sets

$a(a|b)^*$  generates all strings of  $a$ 's and  $b$ 's beginning with an  $a$

**Finite automata** recognize the regular sets



This automaton recognizes the same set of strings.

# Specifying Syntax: Context-free Grammars

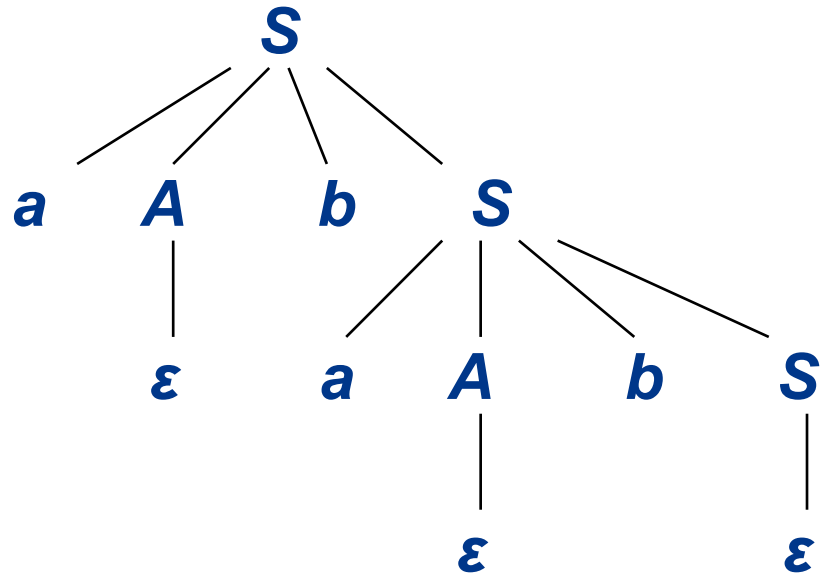
This grammar  $G$  generates **all strings of  $a$ 's and  $b$ 's with the same number of  $a$ 's as  $b$ 's**:

$$S \rightarrow aAbS \mid bBaS \mid \varepsilon$$

$$A \rightarrow aAbA \mid \varepsilon$$

$$B \rightarrow bBaB \mid \varepsilon$$

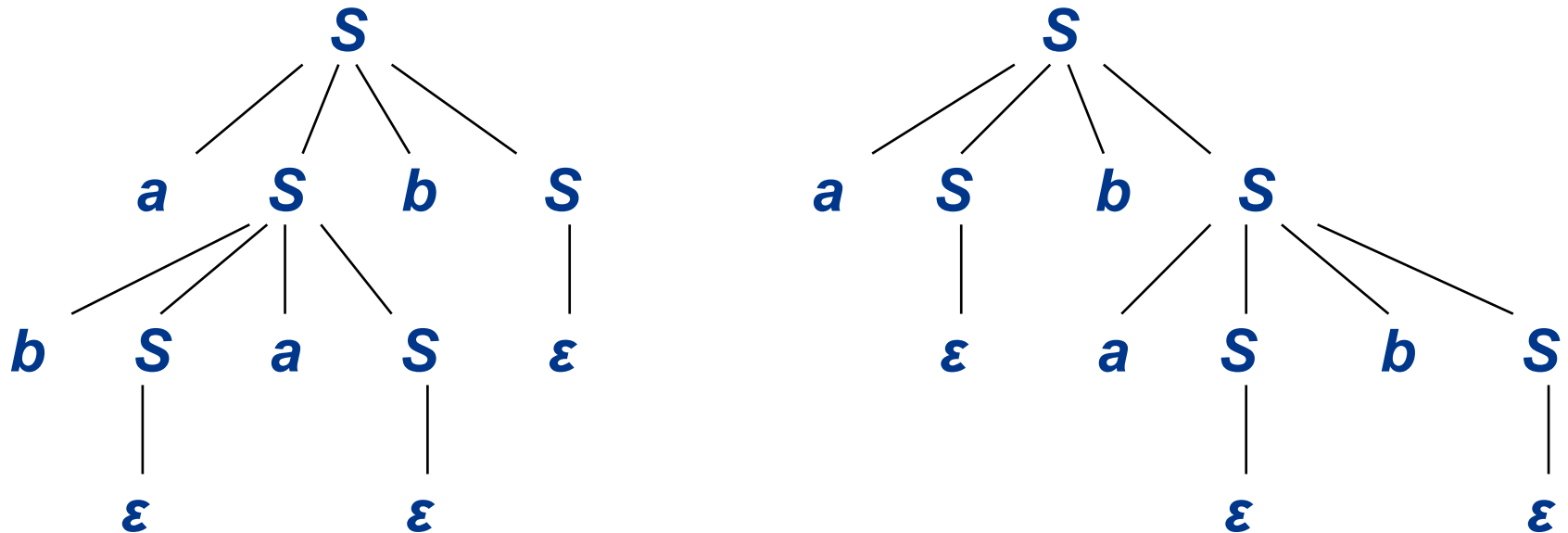
**$G$  is unambiguous and has only one parse tree for every sentence in  $L(G)$ .**



# There is an Art to Writing Good Grammars

The grammar  $S \rightarrow aSbS \mid bSaS \mid \varepsilon$  also generates **all strings of  $a$ 's and  $b$ 's with the same number of  $a$ 's as  $b$ 's**.

But this grammar is **ambiguous**:  $abab$  has two parse trees



$(ab)^n$  has  $\frac{1}{n+1} \binom{2n}{n}$  parse trees

# Natural Languages are Inherently Ambiguous

***I made her duck.***

[5 meanings: D. Jurafsky and J. Martin, 2000]

***One morning I shot an elephant in my pajamas. How he got into my pajamas I don't know.***

[Groucho Marx, *Animal Crackers*, 1930]

***List the sales of the products produced in 1973 with the products produced in 1972.***

[455 parses: W. Martin, K. Church, R. Patil, 1987]