

COMP2012/G52LAC  
Languages and Computation  
Lecture 9  
*The Language of a CFG*

Henrik Nilsson

University of Nottingham

# Recap: Definition of CFG

A CFG  $G = (N, T, P, S)$  where

- $N$  is a finite set of **nonterminals** (or **variables** or **syntactic categories**)
- $T$  is a finite set of **terminals**
- $N \cap T = \emptyset$  (disjoint)
- $P$  is a finite set of **productions** of the form  $A \rightarrow \alpha$  where  $A \in N$  and  $\alpha \in (N \cup T)^*$
- $S \in N$  is the **start symbol**

# Simple Arithmetic Expressions

$SAE = (N = \{E, I, D\}, T = \{+, *, (, ), 0, 1, \dots, 9\}, P, E)$   
where  $P$  is given by:

$$E \rightarrow E + E$$

$$| E * E$$

$$| (E)$$

$$| I$$

$$I \rightarrow DI | D$$

$$D \rightarrow 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$$

**Note:**  $A \rightarrow \alpha | \beta$  shorthand for  $A \rightarrow \alpha, A \rightarrow \beta$ .

# Another Example: Java

The syntax of programming languages is invariably specified by CFGs.

Example: The Java Language Specification, Third Edition. Section 14.5, page 368 gives a CFG for Java statements.

# The Directly Derives Relation (1)

To formally define the language generated by

$$G = (N, T, P, S)$$

we first define a binary relation  $\Rightarrow_G$  on strings over  $N \cup T$ , read “**directly derives in grammar  $G$** ”, being the least relation such that

$$\alpha A \gamma \Rightarrow_G \alpha \beta \gamma$$

whenever  $A \rightarrow \beta$  is a production in  $G$  where  $A \in N$  and  $\alpha, \beta, \gamma \in (N \cup T)^*$ .

# The Directly Derives Relation (2)

When it is clear which grammar  $G$  is involved, we use  $\Rightarrow$  instead of  $\Rightarrow_G$ .

Example: Given the grammar

$$\begin{aligned} S &\rightarrow \epsilon \mid aA \\ A &\rightarrow bS \end{aligned}$$

we have

$$\begin{aligned} S &\Rightarrow \epsilon & aA &\Rightarrow abS \\ S &\Rightarrow aA & SaAaa &\Rightarrow SabSaa \end{aligned}$$

# The Derives Relation (1)

The relation  $\xRightarrow{*}_G$ , read “**derives in grammar  $G$** ”, is the reflexive, transitive closure of  $\xRightarrow{G}$ .

That is,  $\xRightarrow{*}_G$  is the least relation on strings over  $N \cup T$  such that:

# The Derives Relation (1)

The relation  $\xRightarrow{*}_G$ , read “**derives in grammar  $G$** ”, is the reflexive, transitive closure of  $\xRightarrow{G}$ .

That is,  $\xRightarrow{*}_G$  is the least relation on strings over  $N \cup T$  such that:

- $\alpha \xRightarrow{*}_G \beta$  if  $\alpha \xRightarrow{G} \beta$

# The Derives Relation (1)

The relation  $\xRightarrow{*}_G$ , read “**derives in grammar  $G$** ”, is the reflexive, transitive closure of  $\xRightarrow{G}$ .

That is,  $\xRightarrow{*}_G$  is the least relation on strings over  $N \cup T$  such that:

- $\alpha \xRightarrow{*}_G \beta$  if  $\alpha \xRightarrow{G} \beta$

- $\alpha \xRightarrow{*}_G \alpha$  (reflexive)

# The Derives Relation (1)

The relation  $\xRightarrow{*}_G$ , read “**derives in grammar  $G$** ”, is the reflexive, transitive closure of  $\xRightarrow{G}$ .

That is,  $\xRightarrow{*}_G$  is the least relation on strings over  $N \cup T$  such that:

- $\alpha \xRightarrow{*}_G \beta$  if  $\alpha \xRightarrow{G} \beta$

- $\alpha \xRightarrow{*}_G \alpha$  (reflexive)

- $\alpha \xRightarrow{*}_G \beta$  if  $\alpha \xRightarrow{*}_G \gamma \wedge \gamma \xRightarrow{*}_G \beta$  (transitive)

# The Derives Relation (2)

Again, we use  $\Rightarrow^*$  instead of  $\xRightarrow_G^*$  when  $G$  is obvious.

Example: Given the grammar

$$\begin{aligned} S &\rightarrow \epsilon \mid aA \\ A &\rightarrow bS \end{aligned}$$

we have

$$\begin{array}{ll} S \Rightarrow^* \epsilon & S \Rightarrow^* abS \\ S \Rightarrow^* aA & S \Rightarrow^* ababS \\ aA \Rightarrow^* abS & S \Rightarrow^* abab \end{array}$$

# Lang. Generated by a Grammar

The **language generated** by a context-free grammar

$$G = (N, T, P, S)$$

denoted  $L(G)$ , is defined as follows:

$$L(G) = \{w \mid w \in T^* \wedge S \xRightarrow[G]{*} w\}$$

A language  $L$  is a **Context-Free Language** (CFL) iff  $L = L(G)$  for some CFG  $G$ .

A string  $\alpha \in (N \cup T)^*$  is a **sentential form** iff  $S \xRightarrow{*} \alpha$ .