G52MAL
Machines and Their Languages
Lecture 11
Derivation Trees and Ambiguity

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 $A \to \alpha$ where $A \in N$ and $\alpha \in (N \cup T)^*$
- $S \in N$ is the ***start symbol***

Recap: The Directly Derives Relation (1)

To formally define the language generated by

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

we first define a binary relation $\underset{G}{\Rightarrow}$ on strings over $N \cup T$, read "***directly derives in grammar*** G", being the least relation such that

$$\alpha A \gamma \underset{G}{\Rightarrow} \alpha \beta \gamma$$

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

Recap: The Directly Derives Relation (2)

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

Example: Given the grammar

$$\begin{array}{rcl} S & \to & \epsilon \mid aA \\ A & \to & bS \end{array}$$

we have

$$\begin{array}{rclcrcl} S & \Rightarrow & \epsilon & \quad & aA & \Rightarrow & abS \\ S & \Rightarrow & aA & \quad & SaAaa & \Rightarrow & SabSaa \end{array}$$

Recap: The Derives Relation (1)

The relation $\overset{*}{\underset{G}{\Rightarrow}}$, read "***derives in grammar*** G", is the reflexive, transitive closure of $\underset{G}{\Rightarrow}$.

That is, $\overset{*}{\underset{G}{\Rightarrow}}$ is the least relation on strings over $N \cup T$ such that:

- $\alpha \overset{*}{\underset{G}{\Rightarrow}} \beta$ if $\alpha \underset{G}{\Rightarrow} \beta$
- $\alpha \overset{*}{\underset{G}{\Rightarrow}} \alpha$ (reflexive)
- $\alpha \overset{*}{\underset{G}{\Rightarrow}} \beta$ if $\alpha \overset{*}{\underset{G}{\Rightarrow}} \gamma \wedge \gamma \overset{*}{\underset{G}{\Rightarrow}} \beta$ (transitive)

Recap: The Derives Relation (2)

Again, we use $\overset{*}{\Rightarrow}$ instead of $\overset{*}{\underset{G}{\Rightarrow}}$ when G is obvious.

Example: Given the grammar

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

we have

$$\begin{aligned} S &\overset{*}{\Rightarrow} \epsilon & S &\overset{*}{\Rightarrow} abS \\ S &\overset{*}{\Rightarrow} aA & S &\overset{*}{\Rightarrow} ababS \\ aA &\overset{*}{\Rightarrow} abS & S &\overset{*}{\Rightarrow} abab \end{aligned}$$

Recap: 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:

$$L(G) = \{w \mid w \in T^* \ \wedge \ S \overset{*}{\underset{G}{\Rightarrow}} w\}$$

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 $S \overset{*}{\Rightarrow} \alpha$.

Recap: Language Generation: Example

Given the grammar $G = (N = \{S, A\}, T = \{a, b\}, P, S)$ where P are the productions

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

we have

$$\begin{aligned} L(G) &= \{(ab)^i \mid i \geq 0\} \\ &= \{\epsilon, ab, abab, ababab, abababab, \ldots\} \end{aligned}$$

Simple Arithmetic Expressions

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

$$\begin{array}{rcl} E & \rightarrow & E + E \\ & \mid & E * E \\ & \mid & (E) \\ & \mid & I \\ I & \rightarrow & DI \mid D \\ D & \rightarrow & 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \end{array}$$

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