German University in Cairo
Department of Computer Science
Assoc. Prof. Haythem O. Ismail

CSEN 1003 Compiler, Spring Term 2019
Practice Assignment 3
Discussion: 12.02.19 - 17.02.19

Exercise 3-1
CFG's

Give a context-free grammar (CFG) for each of the following languages:

a) $L = \{\mathtt{a}^m\mathtt{b}^n\mathtt{c}^k \mid k = m + n \text{ and } m, n, k \geq 0\}$ over the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$.

Solution:

$$\begin{array}{lcl} S & \rightarrow & \mathtt{a}S\mathtt{c} \mid T \\ T & \rightarrow & \mathtt{b}T\mathtt{c} \mid \varepsilon \end{array}$$

b) $L = \{\mathtt{a}^m\mathtt{b}^n \mid n \neq m\}$ over the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}\}$.

Solution:

$$\begin{array}{lcl} S & \rightarrow & P \mid T \\ P & \rightarrow & \mathtt{a}P\mathtt{b} \mid \mathtt{a}P \mid \mathtt{a} \\ T & \rightarrow & \mathtt{a}T\mathtt{b} \mid T\mathtt{b} \mid \mathtt{b} \end{array}$$

Alternative solution:

$$\begin{array}{lcl} S & \rightarrow & AX \mid XB \\ X & \rightarrow & \mathtt{a}X\mathtt{b} \mid \varepsilon \\ A & \rightarrow & \mathtt{a}A \mid \mathtt{a} \\ B & \rightarrow & \mathtt{b}B \mid \mathtt{b} \end{array}$$

Note: This language does not accept the empty string because it would imply $m = n = 0$.

c) $L = \{w \mid w \text{ is a palindrome }\}$ over the alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. (Note: A palindrome is a string that reads the same backwards as forwards.)

Solution:

$$S \rightarrow \varepsilon \mid \mathtt{a} \mid \mathtt{b} \mid \mathtt{c} \mid \mathtt{a}S\mathtt{a} \mid \mathtt{b}S\mathtt{b} \mid \mathtt{c}S\mathtt{c}$$

Exercise 3-2
Parse trees

Cosider the grammar:

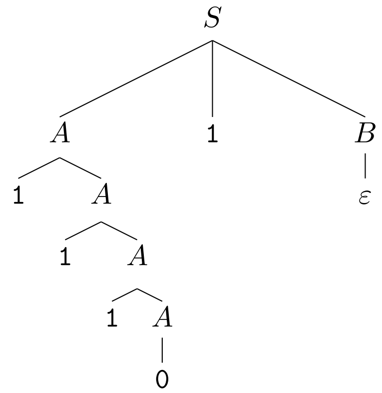
[0]Some exercises are due to Dr. Carmen Gervet

$$
\begin{aligned}
S &\rightarrow A\mathtt{1}B \\
A &\rightarrow \mathtt{1}A \mid \mathtt{0} \\
B &\rightarrow \mathtt{0}B \mid \varepsilon
\end{aligned}
$$

Give a parse tree for each of the following strings:

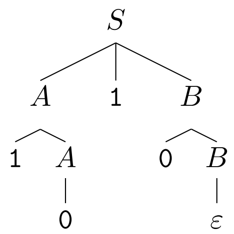
a) `11101`

Solution:



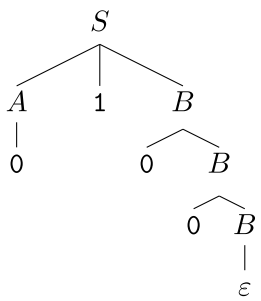
b) `1010`

Solution:



c) `0100`

Solution:



Exercise 3-3

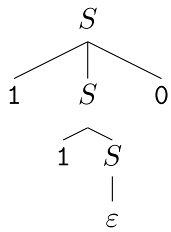
Ambiguous grammars

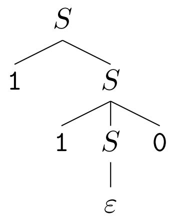
For the following grammars, first show that the grammar is ambiguous, then provide an equivalent unambiguous grammar.

a) $S \quad \rightarrow \quad \mathtt{1}S\mathtt{0} \mid \mathtt{1}S \mid \varepsilon$

Solution:

We show that the grammar is ambiguous by providing two different parse trees for the string:`110`





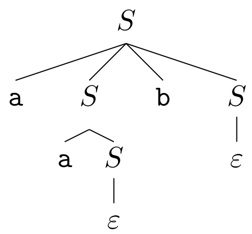
An equivalent unambiguous grammar:

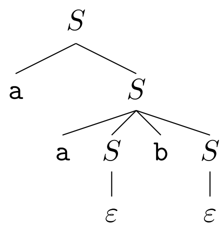
$$
\begin{array}{lcl}
S & \rightarrow & \mathtt{1}S\mathtt{0} \mid T \\
T & \rightarrow & \mathtt{1}T \mid \varepsilon
\end{array}
$$

b) $S \quad \rightarrow \quad \mathtt{a}S\mathtt{b}S \mid \mathtt{a}S \mid \varepsilon$

Solution:

We show that the grammar is ambiguous by providing two different parse trees for the string:`aab`





An equivalent unambiguous grammar:

$$
\begin{array}{lcl}
S & \rightarrow & TS \mid RS \mid A \mid \varepsilon \\
T & \rightarrow & \mathtt{a}T\mathtt{b} \mid \mathtt{ab} \\
R & \rightarrow & \mathtt{a}R\mathtt{b} \mid \mathtt{a}A\mathtt{b} \\
A & \rightarrow & \mathtt{a}A \mid \mathtt{a}
\end{array}
$$

Exercise 3-4

Leftmost and rightmost derivations

Consider the following context-free grammar:

$S \quad \rightarrow \quad SS\mathtt{+} \mid SS\mathtt{*} \mid \mathtt{a}$

and the string: `aa+a*`

a) Give a leftmost derivation for the string. Show the sequence of derivation rules applied.

Solution:

$S \Rightarrow SS\texttt{*} \Rightarrow (SS\texttt{+})S\texttt{*} \Rightarrow \texttt{a}S\texttt{+}S\texttt{*} \Rightarrow \texttt{aa+}S\texttt{*} \Rightarrow \texttt{aa+a*}$

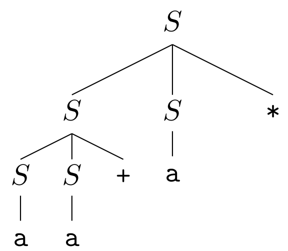
b) Give a rightmost derivation for the string. Show the sequence of derivation rules applied.

Solution:

$S \Rightarrow SS\texttt{*} \Rightarrow S\texttt{a*} \Rightarrow (SS\texttt{+})\texttt{a*} \Rightarrow S\texttt{a+a*} \Rightarrow \texttt{aa+a*}$

c) Give a parse tree for the string.

Solution:



d) Is this grammar ambiguous? Justify your answer.

Solution:

This grammar describes the language of strings in postfix notation with the operand ‘a’. It is not ambiguous because postfix notation implies a single interpretation of strings.

Exercise 3-5

Unambiguous grammars

The following context-free grammar generates prefix expressions with operands 0 and 1 and binary operators +, -, and *:

$S \quad \rightarrow \quad \texttt{+}SS \mid \texttt{-}SS \mid \texttt{*}SS \mid \texttt{0} \mid \texttt{1}$

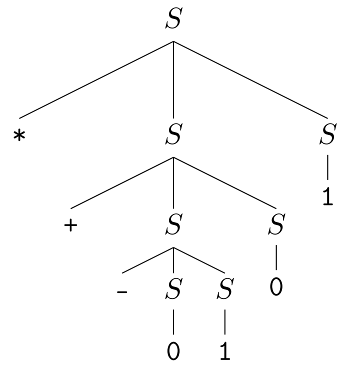
a) Find leftmost and rightmost derivations together with a parse tree for the string *+-0101.

Solution:

Derivations:

Leftmost	**Rightmost**
S	S
$\Rightarrow\texttt{*}SS$	$\Rightarrow\texttt{*}SS$
$\Rightarrow\texttt{*}(\texttt{+}SS)S$	$\Rightarrow\texttt{*}S\texttt{1}$
$\Rightarrow\texttt{*+}(\texttt{-}SS)SS$	$\Rightarrow\texttt{*}(\texttt{+}SS)\texttt{1}$
$\Rightarrow\texttt{*+-0}SSS$	$\Rightarrow\texttt{*+}S\texttt{01}$
$\Rightarrow\texttt{*+-01}SS$	$\Rightarrow\texttt{*+}(\texttt{-}SS)\texttt{01}$
$\Rightarrow\texttt{*+-010}S$	$\Rightarrow\texttt{*+-}S\texttt{101}$
$\Rightarrow\texttt{*+-0101}$	$\Rightarrow\texttt{*+-0101}$

Parse tree:



b) Prove that this grammar is unambiguous.

Solution:

This grammar denotes a prefix notation of strings with operands `0`, `1` and operation symbols `+`, `-` and `*`.

In this grammar, the application of each rule generates a string starting with a unique terminal symbol (`*`, `+`, `-`, `0` or `1`). For any string w that belongs to the CFL, when we consider a leftmost variable E in the leftmost derivation of the string, there is only one rule that can be used to continue the derivation. This rule is uniquely determined by the next symbol in w to be derived. So there is only one leftmost derivation for w, hence the non-ambiguity of the grammar.

Exercise 3-6

Grammar Correctness

a) Consider the CFG G_1:

$$S \longrightarrow \texttt{0}S\texttt{11} \mid \texttt{0}S\texttt{111} \mid \varepsilon$$

Prove that $L(G_1) = \{\texttt{0}^m\texttt{1}^n \mid 2m \leq n \leq 3m \text{ and } n, m \geq 0\}$

Solution:

Proof. We divide the proof into two parts.

Soundness ($\mathbf{L(G_1)} \subseteq \mathbf{L_1}$). We prove the statement by induction on the length k of S-derivations.

Basis ($k = 1$)**.** The only S-derivation of length 1 is the derivation $S \Rightarrow \varepsilon$ and $\varepsilon \in L_1$ when $m = n = 0$.

Induction Hypothesis. For some $k \in \mathtt{N}$ and $\forall j \leq k$, if $S \overset{j}{\Rightarrow} w$, then $w \in L_1$.

Induction Step. Suppose that $S \overset{k+1}{\Rightarrow} w$. Hence either, $S \Rightarrow \texttt{0}S\texttt{11} \overset{k}{\Rightarrow} \texttt{0}u\texttt{11} = w$ or $S \Rightarrow \texttt{0}S\texttt{111} \overset{k}{\Rightarrow} \texttt{0}v\texttt{111} = w$. Thus, $S \overset{k}{\Rightarrow} u$ and $S \overset{k}{\Rightarrow} v$. Then, by the induction hypothesis, $u \in L_1$ and $v \in L_1$.
Hence, $u, v = \texttt{0}^m\texttt{1}^n$, for some $m, n \in \mathtt{N}$ *and* $2m \leq n \leq 3m$. Accordingly, it must be one of three cases:

1. $n = 2m$. In this case, it must be that $w = 0u11 = 0^{m+1}1^{2m+2} \in L_1$; or
2. $n = 3m$. In this case, it must be that $w = 0v111 = 0^{m+1}1^{3m+3} \in L_1$; or

3. $2m < n < 3m$. In this case, either $w = 0u11$ or $w = 0v111$. If $w = 0u11 = 0^{m+1}1^{n+2}$, then $2m+2 < n+2 < 3m+2 < 3m+3$. If $w = 0v111 = 0^{m+1}1^{n+3}$, then $2m+2 < 2m+3 < n+3 < 3m+3$. Hence, in both cases $w \in L_1$.

Completeness ($\mathbf{L_1 \subseteq L(G_1)}$). We prove the statement by induction on the length k of strings in L_1.

Basis ($k = 0$)**.** The only string of length 0 in L_1 is ε, and $S \Rightarrow \varepsilon$. Hence, $\varepsilon \in L(G_1)$.

Induction Hypothesis. For some $k \in \mathtt{N}$, if $|w| \leq k$ and $w \in L_1$, then $w \in L(G_1)$ ($S \overset{*}{\Rightarrow} w$).

Induction Step. Suppose $w \in L_1$ with $|w| = k+1$. By definition of L_1, it must be that $w = \mathtt{0}^m\mathtt{1}^n$, for some $m, n \in \mathtt{N}$ *and* $2m \leq n \leq 3m$. It must be one of three cases:

1. $n = 2m$. Then, $w = 0^m 1^{2m}$. It must be that $w = 0u11$ where $u = 0^{m-1}1^{2m-2}$. Since $|u| \leq k$ and $u \in L_1$, then by the induction hypothesis $S \overset{*}{\Rightarrow} u$. Therefore, a valid derivation for w is $S \Rightarrow 0S11 \overset{*}{\Rightarrow} 0u11 = w$; or
2. $n = 3m$. Then, $w = 0^m 1^{3m}$. It must be that $w = 0v111$ where $v = 0^{m-1}1^{3m-3}$. Since $|v| \leq k$ and $v \in L_1$, then by the induction hypothesis $S \overset{*}{\Rightarrow} v$. Therefore, a valid derivation for w is $S \Rightarrow 0S111 \overset{*}{\Rightarrow} 0v111 = w$; or
3. $2m < n < 3m$. In this case, it must be that w is either $0u11$ where $u = 0^{m-1}1^{n-2}$, or w is $0v111$ where $v = 0^{m-1}1^{n-3}$. It is fairly obvious that $|u|, |v| \leq k$. It only remains to show that $u, v \in L_1$ to use the induction hypothesis. We show this in the following.
   i. Since $2m-2 < n-2 < 3m-2$, then $2m-2 < n-2 \leq 3m-3$. Accordingly, $u \in L_1$. By the induction hypothesis, $S \overset{*}{\Rightarrow} u$. Therefore, a valid derivation for w is $S \Rightarrow 0S11 \overset{*}{\Rightarrow} 0u11 = w$.
   ii. Since $2m-3 < n-3 < 3m-3$, then $2m-2 \leq n-3 < 3m-3$. Accordingly, $v \in L_1$. By the induction hypothesis, $S \overset{*}{\Rightarrow} v$. Therefore, a valid derivation for w is $S \Rightarrow 0S111 \overset{*}{\Rightarrow} 0v111 = w$.

   Thus, $w \in L(G_1)$. □

b) Consider the CFG G_2:

$$\begin{array}{rcl} S & \longrightarrow & AC \\ A & \longrightarrow & \mathtt{a}A\mathtt{b} \mid \varepsilon \\ C & \longrightarrow & \mathtt{c}C \mid \varepsilon \end{array}$$

Prove that $L(G_2) = \{\mathtt{a}^m\mathtt{b}^m\mathtt{c}^n \mid m, n \geq 0\}$

Solution:

First, it should be noted that, since the only S-rule is the rule $S \Rightarrow AC$, every derivation of a string $w \in L(G_2)$ is of the form

$$S \Rightarrow AC \overset{*}{\Rightarrow} uv = w$$

where $A \overset{*}{\Rightarrow} u \in \Sigma^*$ and $C \overset{*}{\Rightarrow} v \in \Sigma^*$. Hence, $L(G_2) = L(G_A) \circ L(G_C) = \{u \mid A \overset{*}{\Rightarrow} u\} \circ \{v \mid C \overset{*}{\Rightarrow} v\}$. To prove that $L(G_2) = \{\mathtt{a}^m\mathtt{b}^m\mathtt{c}^n \mid m, n \geq 0\}$, it suffices to show that

a) $L(G_A) = L_1 = \{\mathtt{a}^m\mathtt{b}^m \mid m \geq 0\}$ and

b) $L(G_C) = L_2 = \{\mathtt{c}^n \mid n \geq 0\}$.

Claim 1. $L(G_A) = L_1$

Proof. We divide the proof into two parts.

$\mathbf{L(G_A) \subseteq L_1}$**.** We prove the statement by induction on the length k of A-derivations.

Basis ($k = 1$)**.** The only A-derivation of length 1 is the derivation $A \Rightarrow \varepsilon$ and $\varepsilon \in L_1$.

Induction Hypothesis. For some $k \in \mathtt{N}$, if $A \overset{j}{\Rightarrow} w$, $\forall j \leq k$, then $w \in L_1$.

Induction Step. Suppose that $A \overset{k+1}{\Rightarrow} w$. Hence,

$$A \Rightarrow \mathtt{a}A\mathtt{b} \overset{k}{\Rightarrow} \mathtt{a}u\mathtt{b} = w$$

Thus, $A \overset{k}{\Rightarrow} u$. By the induction hypothesis, $u \in L_1$. Hence, $u = \mathtt{a}^m\mathtt{b}^m$, for some $m \geq 0$. It follows that $w = \mathtt{a}u\mathtt{b} = \mathtt{a}^{m+1}\mathtt{b}^{m+1} \in L_1$.

$\mathbf{L_1 \subseteq L(G_A)}$**.** We prove the statement by induction on the length k of strings in L_1.

Basis ($k = 0$)**.** The only string of length 0 in L_1 is ε, and $A \Rightarrow \varepsilon$. Hence, $\varepsilon \in L(G_A)$.

Induction Hypothesis. For some $k \in \mathtt{N}$, if $|w| \leq k$ and $w \in L_1$, then $w \in L(G_A)$.

Induction Step. Let $w \in L_1$ with $|w| = k + 1$. By definition of L_1, $w = \mathtt{a}^m\mathtt{b}^m$, for some $m \geq 0$. Moreover, since $|w| = k + 1$, it follows that $m \geq 1$. Hence, $w = \mathtt{a}u\mathtt{b}$, where $u = \mathtt{a}^{m-1}\mathtt{b}^{m-1}$ for some $m - 1 \geq 0$. Thus, $u \in L_1$. Moreover, since $2m = k + 1$, it follows that $|u| = 2m - 2 = k - 1$. Hence, by the induction hypothesis, $u \in L_A$. By definition of $L(G_A)$, $A \overset{*}{\Rightarrow} u$. Thus, the following is a valid A-derivation:

$$A \Rightarrow \mathtt{a}A\mathtt{b} \overset{*}{\Rightarrow} \mathtt{a}u\mathtt{b} = w$$

Thus, $w \in L(G_A)$. □

Claim 2. $L(G_C) = L_2$

Proof. We divide the proof into two parts.

$\mathbf{L(G_C) \subseteq L_2}$**.** We prove the statement by induction on the length k of C-derivations.

Basis ($k = 1$)**.** The only C-derivation of length 1 is the derivation $C \Rightarrow \varepsilon$ and $\varepsilon \in L_2$.

Induction Hypothesis. For some $k \in \mathtt{N}$, if $C \overset{j}{\Rightarrow} w$, $\forall j \leq k$, then $w \in L_2$.

Induction Step. Suppose that $C \overset{k+1}{\Rightarrow} w$. Hence,

$$C \Rightarrow \mathtt{c}C \overset{k}{\Rightarrow} \mathtt{c}u = w$$

Thus, $C \overset{k}{\Rightarrow} u$. By the induction hypothesis, $u \in L_2$. Hence, $u = \mathtt{c}^n$, for some $n \geq 0$. It follows that $w = \mathtt{c}u = \mathtt{c}^{n+1} \in L_2$.

$\mathbf{L_2 \subseteq L(G_C)}$**.** We prove the statement by induction on the length k of strings in L_2.

Basis ($k = 0$)**.** The only string of length 0 in L_2 is ε, and $C \Rightarrow \varepsilon$. Hence, $\varepsilon \in L(G_C)$.

Induction Hypothesis. For some $k \in \mathtt{N}$, if $|w| \leq k$ and $w \in L_2$, then $w \in L(G_C)$.

Induction Step. Let $w \in L_2$ with $|w| = k + 1$. By definition of L_2, $w = \mathtt{c}^n$, for some $n \geq 0$. Moreover, since $|w| = k + 1$, it follows that $n \geq 1$. Hence, $w = \mathtt{c}u$, where $u = \mathtt{c}^{n-1}$ for some $n - 1 \geq 0$. Thus, $u \in L_2$. Moreover, since $n = k + 1$, it follows that $|u| = n - 1 = k$. Hence, by the induction hypothesis, $u \in L(G_C)$. By definition of $L(G_C)$, $C \overset{*}{\Rightarrow} u$. Thus, the following is a valid C-derivation:

$$C \Rightarrow \mathtt{c}C \overset{*}{\Rightarrow} \mathtt{c}u = w$$

Thus, $w \in L(G_C)$. □