

Compiler Construction 2016

— Solution Series 3 —

Hand in until May 24th before the exercise class.



[audience looks around] “What just happened?”, “There must be some context we’re missing”

Credit: <https://xkcd.com/1090/>

General Remarks

- Follow the naming convention for the zip file: `ex2_MATRNO1_MATRNO2_MATRNO3` and include the complete framework provided to you via our webpage.
- It is allowed to hand your solutions for the theoretical part via email **as a separately attached PDF file**.
- Please write your name and matriculation number on your solution.

Exercise 1

(2 Points)

Consider the context-free grammar G given by the following rules:

$$S \rightarrow S + S \mid S S \mid (S) \mid S^* \mid a$$

- Provide a leftmost analysis of the string $(a + a)^*a$.
- Provide a rightmost analysis of the string $(a + a)^*a$.
- Prove or disprove: G is unambiguous.

Solution: _____

We enumerate the rules from left to right starting with 1.

(a)

$$\begin{aligned}
 & S \\
 \Rightarrow_l^2 & S S \\
 \Rightarrow_l^4 & S^* S \\
 \Rightarrow_l^3 & (S)^* S \\
 \Rightarrow_l^1 & (S + S)^* S \\
 \Rightarrow_l^5 & (a + S)^* S \\
 \Rightarrow_l^5 & (a + a)^* S \\
 \Rightarrow_l^5 & (a + a)^* a
 \end{aligned}$$

Hence, a leftmost analysis of $(a + a)^* a$ is $z_l = 2431555$.

(b)

$$\begin{aligned}
 & S \\
 \Rightarrow_r^2 & S S \\
 \Rightarrow_r^5 & S a \\
 \Rightarrow_r^4 & S^* a \\
 \Rightarrow_r^3 & (S)^* a \\
 \Rightarrow_r^1 & (S + S)^* a \\
 \Rightarrow_r^5 & (S + a)^* a \\
 \Rightarrow_r^5 & (a + a)^* a
 \end{aligned}$$

Hence, a rightmost analysis of $(a + a)^* a$ is $z_r = 2543155$.

(c)

G is ambiguous. For example, there exist two leftmost analysis for the word aaa , namely $z_1 = 22555$ and $z_2 = 25255$.

Exercise 2

(2 Points)

Complete the correctness proof of Theorem 6.1 by showing the direction omitted in the lecture. More precisely, let $G = \langle N, \Sigma, P, S \rangle \in \text{CFG}_\Sigma$ and $NTA(G)$ as in the lecture (lecture 6, slide 5). Show that for each $w \in \Sigma^*$ and $z \in [p]^*$ it holds that

$$(w, S, \varepsilon) \vdash^* (\varepsilon, \varepsilon, z) \quad \text{implies} \quad z \text{ is a leftmost analysis of } w.$$

Solution: _____

More general: $(w, \alpha, y) \vdash^* (\varepsilon, \varepsilon, yz)$ implies $\alpha \rightarrow_l^z w$

Induction over $n = |z|$:

$n = 0$: then $\alpha = w$ as there may only be matching steps and thus

$$(w, w, y) \vdash^* (\varepsilon, \varepsilon, y) \text{ implies } w \rightarrow_l^\varepsilon w$$

$n \rightarrow n + 1$: for $z = iz'$ we know from induction that if $(w', \alpha', y') \vdash^* (\varepsilon, \varepsilon, y'z')$ then $\alpha' \rightarrow_l^{z'} w'$

Let $\alpha = uA\beta$, $w = uv$, where $u, v \in \Sigma^*$ and let $\pi(i) = A \rightarrow \gamma$, then

$$(w, \alpha, y) = (uv, uA\beta, y) \vdash^* (v, A\beta, y) \vdash (v, \gamma\beta, \underline{yi}) \xrightarrow{\text{ind.hyp.}} \vdash^* (\varepsilon, \varepsilon, \underline{yiz'})$$

The above implies:

$$\begin{array}{llll} \gamma\beta & \rightarrow_l^{z'} & v & \text{(from induction hypothesis)} \\ u\gamma\beta & \rightarrow_l^{z'} & uv = w & \\ \alpha = uA\beta & \rightarrow_l^i & u\gamma\beta & \rightarrow_l^{z'} \quad uv = w \quad (\text{as } \pi(i) = A \rightarrow \gamma) \\ \alpha = uA\beta & \rightarrow_l^{iz'=z} & uv = w & \end{array}$$

Exercise 3

(1 Points)

Show that for each context free grammar G it holds that $G \in LL(1)$ implies that G is unambiguous.

Solution: _____

We have $G \in LL(1)$ implies G unambiguous if and only if G ambiguous implies $G \notin LL(1)$.

Now, assume $G = (N, \Sigma, P, S)$ is ambiguous.

$$\Rightarrow \exists \omega \in \Sigma^* \text{ with at least two left-derivations } S \Rightarrow_l^* vA\alpha \begin{cases} \Rightarrow_l v\beta\alpha \Rightarrow_l^* vv' = \omega \\ \Rightarrow_l v\gamma\alpha \Rightarrow_l^* vv' = \omega \end{cases}$$

We assume this splitting to be the first difference appearing in the derivations.

As obviously $fi(v') = fi(v')$ and therefore $fi(\beta\alpha) \cap fi(\gamma\alpha) \neq \emptyset$

$$\Rightarrow G \notin LL(1)$$

Exercise 4

(2 Points)

Two characterizations of $LL(1)$ have been given in the lecture.

First, by Lemma 6.5, a context free grammar $G = \langle N, \Sigma, P, S \rangle$ is in $LL(1)$ if and only if for all leftmost derivations of the form

$$S \Rightarrow_l^* wA\alpha \begin{cases} \Rightarrow_l w\beta\alpha \\ \Rightarrow_l w\gamma\alpha \end{cases}$$

such that $\beta \neq \gamma$, it follows that $first_1(\beta\alpha) \cap first_1(\gamma\alpha) = \emptyset$.

Second, by Theorem 6.10, G is in $LL(1)$ if and only if for all pairs of rules $A \rightarrow \beta | \gamma \in P$, where $\beta \neq \gamma$, we have

$$la(A \rightarrow \beta) \cap la(A \rightarrow \gamma) = \emptyset.$$

- (a) Lift the statement of Theorem 6.10 to $LL(k)$ for arbitrary $k \in \mathbb{N}_{>0}$.
- (b) Show that the two characterizations are not equivalent for $k > 1$ by showing that the following grammar is in $LL(2)$ according to the first characterization, but not according to your modified version of the second one.

$$\begin{aligned} S &\rightarrow aAab \mid bAbb \\ A &\rightarrow a \mid \varepsilon \end{aligned}$$

Solution: _____

(a)

$G \in LL(k)$ iff for all pairs of rules $A \rightarrow \beta \mid \gamma \in P$ (where $\beta \neq \gamma$):

$$la_k(A \rightarrow \beta) \cap la_k(A \rightarrow \gamma) = \emptyset$$

where $la_k(A \rightarrow \beta) = \text{first}_k(\beta \text{ follow}_k(A))$.

(b)

$$\begin{aligned} \text{first}_2(S) &= \{aa, ba, bb\} \\ \text{first}_2(A) &= \{a, \varepsilon\} \\ \text{follow}_2(S) &= \{\varepsilon\} \\ \text{follow}_2(A) &= \{ab, bb\} \end{aligned}$$

$$\begin{aligned} la_2(S \rightarrow aAab) &= \{aa\} \\ la_2(S \rightarrow bAbb) &= \{ba, bb\} \\ la_2(A \rightarrow a) &= \{aa, ab\} \\ la_2(A \rightarrow \varepsilon) &= \{ab, bb\} \end{aligned}$$

$la_2(A \rightarrow a) \cap la_2(A \rightarrow \varepsilon) = \{ab\} \neq \emptyset$ and therefore the grammar is not in $LL(2)$ according to the second definition.

Check all leftmost derivations as in the definition:

$$\begin{aligned} S &\rightarrow_l aAab \\ S &\rightarrow_l bAbb \end{aligned}$$

$$\text{first}_2(aAab) \cap \text{first}_2(bAbb) = \{aa\} \cap \{ba, bb\} = \emptyset$$

$$\begin{aligned} S &\rightarrow_l aAab \rightarrow_l aaab \\ S &\rightarrow_l aAab \rightarrow_l aab \end{aligned}$$

$$\text{first}_2(aab) \cap \text{first}_2(ab) = \{aa\} \cap \{ab\} = \emptyset$$

$$\begin{aligned} S &\rightarrow_l b Abb \rightarrow_l b abb \\ S &\rightarrow_l b Abb \rightarrow_l b bb \end{aligned}$$

$$\text{first}_2(abb) \cap \text{first}_2(bb) = \{ab\} \cap \{bb\} = \emptyset$$

Thus, the grammar is in $LL(2)$ according to the first definition.

Exercise 5

(3 Points)

Consider the following grammar G :

$$\begin{aligned} S &\rightarrow (L) \mid a \\ L &\rightarrow L, S \mid L, SS \mid S \mid SS \end{aligned}$$

- (a) Show that $G \notin LL(1)$.
- (b) Transform G into an equivalent grammar $LL(1)$, i.e. provide a grammar $G' \in LL(1)$ such that $L(G') = L(G)$.
- (c) Prove that G' has the $LL(1)$ property

Solution: _____

(a)

First of $\text{la}(L \rightarrow L, S) \cap \text{La}(L \rightarrow L, SS)$ is not empty.

(b)

First remove left recursion:

$$\begin{aligned} S &\rightarrow (L) \mid a \\ L &\rightarrow SL' \mid SSL' \\ L' &\rightarrow \epsilon, SL' \mid SSL' \mid \epsilon \end{aligned}$$

Then apply left factorisation:

$$\begin{aligned} S &\rightarrow (L) \mid a \\ L &\rightarrow SL'' \\ L' &\rightarrow \epsilon, SL'' \mid \epsilon \\ L'' &\rightarrow L' \mid SL' \end{aligned}$$

(c)

Rule	la
$S \rightarrow (L)$	$\{(\}$
$S \rightarrow a$	$\{a\}$
$L' \rightarrow, SL''$	$\{, \}$
$L' \rightarrow \varepsilon$	$\{\}$
$L'' \rightarrow L'$	$\{, \}$
$L'' \rightarrow SL'$	$\{(a\}$

For every nonterminal it holds that the intersection of their lookahead sets is empty. Therefore the grammar is $LL(1)$.