Parsing
Exercises for mid term exam

Laura Kallmeyer

Winter 2016, Heinrich-Heine-Universität Düsseldorf

Question 1 (CFG) *Consider the following language:*

$$L_1 = \{a^n(bc)^m d^n \mid n \geq 0, m > 0\}$$

1. *Does* $\varepsilon \in L_1$ *hold?*
2. *Give all words of length* 6 *that are in* L_1.
3. *Give a CFG that generates* L_1.

   (It is sufficient to give the productions if all upper case letters are taken to be nonterminals and all lower case letters terminals, and S *is the start symbol.)*

Solution:

1. $\varepsilon \notin L(G)$.
2. *bcbcbc*, *abcbcd*, *aabcdd*
3. $S \rightarrow aSd \mid T$, $T \rightarrow bc \mid bcT$

Question 2 (PDA) *Consider the following PDA* M:

$M = \langle \{q_0, q_1, q_2\}, \{a, b, c, d\}, \{S, T\}, \delta, q_0, S, \emptyset\rangle$ *with*

$\delta(q_0, a, S) = \{\langle q_0, S\rangle\}$ $\quad \delta(q_0, b, S) = \{\langle q_1, T\rangle\}$
$\delta(q_1, c, T) = \{\langle q_0, T\rangle\}$ $\quad \delta(q_0, b, T) = \{\langle q_1, TT\rangle\}$
$\delta(q_0, d, T) = \{\langle q_2, \varepsilon\rangle\}$ $\quad \delta(q_2, d, T) = \{\langle q_2, \varepsilon\rangle\}$

The acceptance is with empty stack, i.e., we consider $N(M)$.

1. *Does* $\varepsilon \in N(M)$ *hold?*
2. *Give all configurations (triple of state, stack and remaining input) that the automaton goes through when processing the input abcd.*
3. *What is the language accepted with empty stack, i.e., what is* $N(M)$*?*

Solution:

1. $\varepsilon \notin N(M)$.
2.

state	stack	rem. input
q_0	S	*abcd*
q_0	S	*bcd*
q_1	T	*cd*
q_0	T	*d*
q_2	ε	ε

3. $\{a^n(bc)^m d^m \mid n \geq 0, m > 0\}$

Question 3 (Unger with deduction rules) *Consider the CFG G with $N = \{S\}, T = \{a\}$, start symbol S and productions*

$$S \to SS \,|\, a$$

Give all the items that arise when doing Unger parsing with deduction rules for the input $w = aaa$.[1]

List the items in a table, giving the antecedent items and the operation an item was created with. In case of several possibilities to generate an item, one of the possibilities is sufficient.

Nr.	*Item*	*Operation/antecedent items*
1	$[\bullet S, 0, 3]$	*Axiom*
2	$[\bullet S, 0, 1]$	*predict from 1*
	...	

Solution:

Nr.	Item	Operation/Antezedensitems
1.	$[\bullet S, 0, 3]$	Axiom
2.	$[\bullet S, 0, 1]$	predict from 1.
3.	$[\bullet S, 1, 3]$	predict from 1.
4.	$[\bullet S, 0, 2]$	predict from 1.
5.	$[\bullet S, 2, 3]$	predict from 1.
6.	$[\bullet a, 0, 1]$	predict from 2.
7.	$[\bullet S, 1, 2]$	predict from 3.
8.	$[\bullet a, 1, 2]$	predict from 7.
9.	$[\bullet a, 2, 3]$	predict from 5.
10.	$[a\bullet, 0, 1]$	scan from 6.
11.	$[a\bullet, 1, 2]$	scan from 8.
12.	$[a\bullet, 2, 3]$	scan from 9.
13.	$[S\bullet, 0, 1]$	complete from 10. and 2.
14.	$[S\bullet, 1, 2]$	complete from 11. and 7.
15.	$[S\bullet, 2, 3]$	complete from 12. and 5.
16.	$[S\bullet, 0, 2]$	complete from 13, 14 and 4
17.	$[S\bullet, 1, 3]$	complete from 14,15 and 3.
18.	$[S\bullet, 0, 3]$	complete from 13, 17 and 1.

Question 4 (Unger deduction rules, MA question)

Modify the Unger deduction rules for the special case of Chomsky normal form grammars.

Hint:
1. You can already perform scan when predicting an A with a matching production $A \to w_{i+1}$.
2. Predict and Complete can be simplified due to the reduced number of righthand side symbols in the productions.

Solution:

Axiom: $\dfrac{}{[\bullet S, 0, n]} \quad |w| = n$

Scan: $\dfrac{[\bullet A, i, i+1]}{[A\bullet, i, i+1]} \quad A \to w_{i+1} \in P$

[1] Axiom: $\dfrac{}{[\bullet S, 0, n]} \quad |w| = n$ Scan: $\dfrac{[\bullet a, i, i+1]}{[a\bullet, i, i+1]} \quad w_{i+1} = a$

Predict: $\dfrac{[\bullet A, i_0, i_k]}{[\bullet A_1, i_0, i_1], \dots, [\bullet A_k, i_{k-1}, i_k]} \quad \begin{array}{l} A \to A_1 \dots A_k \in P, i_j < i_{j+1} \\ \text{if } A_j \in T \text{ then } i_j = i_{j-1} + 1 \\ \text{and } w_j = A_j \end{array}$

Complete: $\dfrac{[\bullet A, i_0, i_k], [A_1\bullet, i_0, i_1], \dots, [A_k\bullet, i_{k-1}, i_k]}{[A\bullet, i_0, i_k]} \quad A \to A_1 \dots A_k \in P$

Predict: $\dfrac{[\bullet A, i, k]}{[\bullet B, i, j], [\bullet C, j, k]} \quad A \to BC \in P, i < j < k$

Complete: $\dfrac{[\bullet A, i, k], [B\bullet, i, j], [C\bullet, j, k]}{[A\bullet, i, k]} \quad A \to BC \in P$

Question 5 (Top-Down parsing) *Consider again the grammar form the preceding questiond (productions $S \to SS \mid a$) and the input $w = aaa$.*

1. *Give all triples of analysis stack, prediction stack and remaining input, that arise in a top-down parsing for this input. Order them in a decision tree that shows which triples arose out of which other triples. Assume as a filter that only triples are licensed where the prediction stack is not longer than the remaining input. (The grammar is ε-free, therefore this assumption.)*

   $\varepsilon \mid S \mid aaa$
   - $\dots$
   - $\dots$

2. *What are the leftmost derivation you can read off the decision tree for aaa?*

Solution:

1. $\varepsilon \mid S \mid aaa$
   - $S_1 \mid SS \mid aaa$
     - $S_1S_1 \mid SSS \mid aaa$
       - $S_1S_1S_2 \mid aSS \mid aaa$
         - $S_1S_1S_2a \mid SS \mid aa$
           - $S_1S_1S_2aS_2 \mid aS \mid aa$
             - $S_1S_1S_2aS_2a \mid S \mid a$
               - $S_1S_1S_2aS_2aS_2 \mid a \mid a$
                 - $S_1S_1S_2aS_2aS_2a \mid \varepsilon \mid \varepsilon$
     - $S_1S_2 \mid aS \mid aaa$
       - $S_1S_2a \mid S \mid aa$
         - $S_1S_2aS_1 \mid SS \mid aa$
           - $S_1S_2aS_1S_2 \mid aS \mid aa$
             - $S_1S_2aS_1S_2a \mid S \mid a$
               - $S_1S_2aS_1S_2aS_2 \mid a \mid a$
                 - $S_1S_2aS_1S_2aS_2a \mid \varepsilon \mid \varepsilon$
         - $S_1S_2aS_2 \mid a \mid aa$
           - $S_1S_2aS_2a \mid \varepsilon \mid a$
   - $S_2 \mid a \mid aaa$
     - $S_2a \mid \varepsilon \mid aa$

2. $S \Rightarrow SS \Rightarrow SSS \Rightarrow aSS \Rightarrow aaS \Rightarrow aaa$
   $S \Rightarrow SS \Rightarrow aS \Rightarrow aSS \Rightarrow aaS \Rightarrow aaa$

Question 6 (CYK-Parsing) *Consider again the CFG from above with productions $S \to SS \mid a$.*

1. *Use the CYK parser for CFGs in Chomsky normal form. Use the version that writes entire productions into the chart, annotated with indices.*

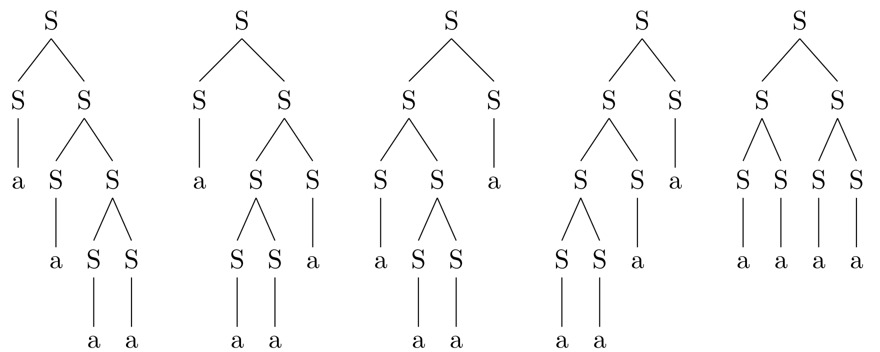
   How does the chart look like that we obtain for an input $w = aaaa$?

2. *Which parse trees for aaaa can be read off the chart?*

Solution:

1. Chart:

l					
4	$S \to S_{1,1}S_{2,3},\ S \to S_{1,2}S_{3,2},\ S \to S_{1,3}S_{4,1}$				
3	$S \to S_{1,1}S_{2,2},\ S \to S_{1,2}S_{3,1}$	$S \to S_{2,1}S_{3,2},\ S \to S_{2,2}S_{4,1}$			
2	$S \to S_{1,1}S_{2,1}$	$S \to S_{2,1}S_{3,1}$	$S \to S_{3,1}S_{4,1}$		
1	$S \to a$	$S \to a$	$S \to a$	$S \to a$	
	1	2	3	4	i

2. 

Question 7 (LL(1) grammars)

1. *Is the grammar from the preceding questions (productions $S \to SS \mid a$) a LL(1) grammar? Explain your answer.*
2. *Consider a CFG with non-terminals $\{S, T\}$, terminals $\{a\}$, start symbol S and productions*

$$S \to TS \mid \varepsilon, T \to a$$

This CFG generates the same language as the one in 1., but with different parse trees. Is this CFG a LL(1) grammar? Explain your answer.

Solution:

1. No, since $Fist(SS) = First(a) = \{a\}$.
2. Yes. The LL(1) parse table is as follows:

	S	T
a	$S \to TS$	$T \to a$
\$	$S \to \varepsilon$	–

There is at most one production per field, consequently this grammar is LL(1).