# Parsing
# Mid term exam, 06.12.2016

Laura Kallmeyer

Winter 2016, Heinrich-Heine-Universität Düsseldorf

Klausurdauer: 90 Minuten.

Hilfmittel: Sämtliche Unterrichtsmaterialien und Notizen in nicht-elektronischer Form.

Questions can be answered in English or in German.

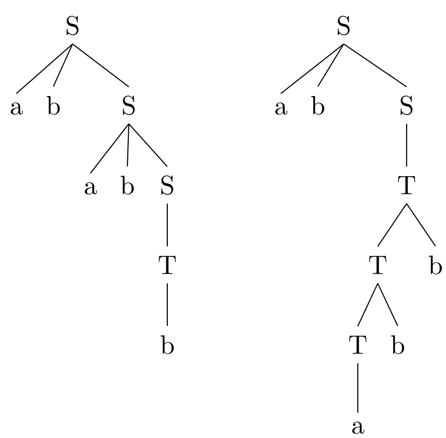
Exercises marked "BA" are only for BA students (APs or BNs), exercises marked "MA" only for MA students. All other exercises are for both.

**Question 1 (BA CFG, 10 pts)**

1. *Give a CFG that generates the language* $\{a^n c^m b^n \mid n > 0, m \geq 0\}$.
2. *Consider the CFG* $G$ *with non-terminals* $N = \{S, T\}$*, terminals* $T = \{a, b\}$*, start symbol* $S$ *and productions* $S \to abS \mid T, T \to Tb \mid a \mid b$.
   (a) *Does* $\varepsilon \in L(G)$ *hold?*
   (b) *Give the two parse trees for ababb that one obtains with* $G$.
   (c) *What is the language* $L(G)$ *generated by this grammar?*

Solution:

1. $S \to aSb \mid aTb, T \to cT \mid \varepsilon$ 3 pts
2. (a) $\varepsilon \notin L(G)$. 1 pt
   (b) 2 pts




   (c) The language can be characterized by a regular expression: $(ab)^*(a|b)b^*$ 4 pts

**Question 1 (MA CFG, 10 pts)**

1. *Give a CFG that generates the language* $\{a^n b^m \mid n \geq m \geq 1\}$.
2. *Consider the CFG* $G$ *with non-terminals* $N = \{S\}$*, terminals* $T = \{a, b\}$*, start symbol* $S$ *and productions* $S \to aSb \mid aSbb \mid \varepsilon$.
   (a) *Does* $\varepsilon \in L(G)$ *hold?*

(b) *Give the two parse trees for aabbb that one obtains with G.*

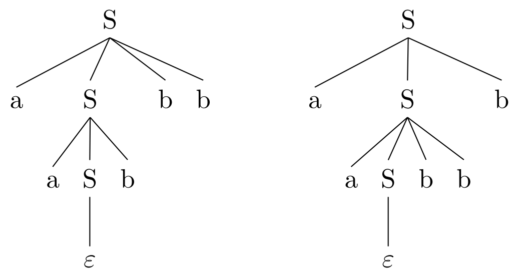
(c) *What is the language $L(G)$ generated by this grammar?*

Solution:

1. $S \to ab \,|\, aS \,|\, aSb$ — 3 pts

2. (a) $\varepsilon \in L(G)$. — 1 pt

   (b) — 2 pts




   (c) $\{a^n b^m \,|\, 2n \geq m \geq n \geq 0\}$ — 4 pts

**Question 2 (PDA, 6 pts)**

*Consider the following PDA M:*

$M = \langle \{q_0, q_1\}, \{a, c\}, \{\#, A\}, \delta, q_0, \#, \emptyset \rangle$ *with*

$\delta(q_0, a, \#) = \{\langle q_0, AA \rangle\}$ $\quad \delta(q_0, a, A) = \{\langle q_0, AAA \rangle\}$

$\delta(q_0, c, A) = \{\langle q_1, A \rangle\}$ $\quad \delta(q_1, a, A) = \{\langle q_1, \varepsilon \rangle\}$

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

1. *Give all configurations (triple of state, stack and remaining input) that the automaton goes through when processing the input aacaaaa.*

2. *What is the language accepted with empty stack, i.e., what is $N(M)$?*

Solution:

1. (3 pts)

| state | stack | rem. input |
|---|---|---|
| $q_0$ | $\#$ | $aacaaaa$ |
| $q_0$ | $AA$ | $acaaaa$ |
| $q_0$ | $AAAA$ | $caaaa$ |
| $q_1$ | $AAAA$ | $aaaa$ |
| $q_1$ | $AAA$ | $aaa$ |
| $q_1$ | $AA$ | $aa$ |
| $q_1$ | $A$ | $a$ |
| $q_1$ | $\varepsilon$ | $\varepsilon$ |

2. $\{a^n c a^{2n} \,|\, n \geq 1\}$ — (3 pts)

**Question 3 (Unger with deduction rules (7 pts))** *Consider the version of the Unger parsing for CNF, formulated with deduction rules:*

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

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

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

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

*Furthermore, take the CFG $G$ with $N = \{S, A\}, T = \{a, b, c\}$, start symbol $S$ and productions*

$$S \to SS \mid SA \mid a, A \to b \mid c$$

*Consider an agenda-based chart parsing of the input word aac, using the CNF Unger chart parser. Give the items that are generated in a table, showing each time the new agenda and the new items generated from the item that has just been removed from the agenda (possibly in combination with other chart items).*

*In other words, from one line to the next 1. remove the first item from the agenda and 2. compute all new items that you can generate from this item and other chart items. These new items are listed in the left column and are appended to the agenda.*

| *new chart items* | *agenda items* |
|---|---|
| $[\bullet S, 0, 3]$ | $[\bullet S, 0, 3]$ |
| $[\bullet S, 0, 1]$ $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ | $[\bullet S, 0, 1]$ $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ |
| $[S\bullet, 0, 1]$ | $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ |
| ... | ... |

Solution:

| new chart items | agenda items |
|---|---|
| $[\bullet S, 0, 3]$ | $[\bullet S, 0, 3]$ |
| $[\bullet S, 0, 1]$ $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ | $[\bullet S, 0, 1]$ $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ |
| $[S\bullet, 0, 1]$ | $[\bullet S, 0, 2]$ $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ |
| $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ | $[\bullet S, 1, 3]$ $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ |
| – | $[\bullet S, 2, 3]$ $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ |
| – | $[\bullet A, 1, 3]$ $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ |
| – | $[\bullet A, 2, 3]$ $[S\bullet, 0, 1]$ $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ |
| $[A\bullet, 2, 3]$ | $[S\bullet, 0, 1]$ $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ $[A\bullet, 2, 3]$ |
| – | $[\bullet S, 1, 2]$ $[\bullet A, 1, 2]$ $[A\bullet, 2, 3]$ |
| $[S\bullet, 1, 2]$ | $[\bullet A, 1, 2]$ $[A\bullet, 2, 3]$ $[S\bullet, 1, 2]$ |
| – | $[A\bullet, 2, 3]$ $[S\bullet, 1, 2]$ |
| $[S\bullet, 1, 3]$ | $[S\bullet, 1, 2]$ $[S\bullet, 1, 3]$ |
| $[S\bullet, 0, 2]$ | $[S\bullet, 1, 3]$ $[S\bullet, 0, 2]$ |
| $[S\bullet, 0, 3]$ | $[S\bullet, 0, 2]$ $[S\bullet, 0, 3]$ |
| – | $[S\bullet, 0, 3]$ |
| – | – |

(7 points)

**Question 4 (Top-Down parsing (6 pts))** *Consider the CFG $G$ with $N = \{S, A\}, T = \{a, c\}$, start symbol $S$ and productions*

$$S \to SS \mid SA \mid a, A \to c$$

*and the input $w = ac$.*

1. *What is the leftmost derivation for ac given this grammar?*

2. *Give all pairs of prediction stack and remaining input that arise in a top-down parsing for this input. List them in a table with a unique number for each pair and indicate from which other pair and with which operation (in particular with which predicted production) a new pair was obtained.*

   *Assume that we do not generate pairs where the stack is longer than the remaining input.*

| *id* | *stack* | *rem. input* | *operation* |
|---|---|---|---|
| *1.* | *S* | *ac* | *axiom* |
| *2.* | *SS* | *ac* | *predict from 1. with* $S \to SS$ |
| *3.* | ... | ... | |

Solution:

1. $S \Rightarrow SA \Rightarrow aA \Rightarrow ac$ — 1 pt

2. (5 pts)

| id | stack | rem. input | operation |
|---|---|---|---|
| 1. | S | ac | axiom |
| 2. | SS | ac | predict from 1. with $S \to SS$ |
| 3. | SA | ac | predict from 1. with $S \to SA$ |
| 4. | a | ac | predict from 1. with $S \to a$ |
| 5. | aS | ac | predict from 2. with $S \to a$ |
| 6. | aA | ac | predict from 3. with $S \to a$ |
| 7. | $\varepsilon$ | c | scan from 4. |
| 8. | S | c | scan from 5. |
| 9. | A | c | scan from 6. |
| 10. | a | c | predict from 8. with $S \to a$ |
| 11. | c | c | predict from 9. with $A \to c$ |
| 12. | $\varepsilon$ | $\varepsilon$ | scan from 11. |

**Question 5 (Shift-reduce parsing (4 pts))** *Consider again the grammar and input from the preceding question.*

1. *What is the rightmost derivation for ac given this grammar?*

2. *Give all pairs of stack and remaining input that arise in a shift-reduce parsing for this input. List them in a table with a unique number for each pair and indicate from which other pair and with which operation (in particular with which reduce production) a new pair was obtained.*

   *Assume that whenever we have a terminal on top of the stack, we perform only a reduce operation (since terminals appear only in righthand sides of length 1).*

| *id* | *stack* | *rem. input* | *operation* |
|---|---|---|---|
| *1.* | $\varepsilon$ | *ac* | *axiom* |
| *2.* | *a* | *c* | *shift from 1.* |
| *3.* | *S* | *c* | *reduce from 2. with* $S \to a$ |
| | ... | ... | |

Solution:

1. $S \Rightarrow SA \Rightarrow Sc \Rightarrow ac$ — 1 pt

2. (3 pts)

| id | stack | rem. input | operation |
|---|---|---|---|
| 1. | $\varepsilon$ | ac | axiom |
| 2. | a | c | shift from 1. |
| 3. | S | c | reduce from 2. with $S \to a$ |
| 4. | Sc | $\varepsilon$ | shift from 3. |
| 5. | SA | $\varepsilon$ | reduce from 4. with $A \to c$ |
| 6. | S | $\varepsilon$ | reduce from 5. with $S \to SA$ |

**Question 6 (CYK-Parsing (7 pts))** *Consider again the CFG from above with productions* $S \to SS \mid SA \mid a, A \to c$.

*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* $w = aacac$*?*

Solution:

Chart:

| $l$ | | | | | | |
|---|---|---|---|---|---|---|
| 5 | $S \to S_{1,1}S_{2,4}$, $S \to S_{1,3}S_{4,2}$, $S \to S_{1,4}A_{5,1}$ | | | | | |
| 4 | $S \to S_{1,1}S_{2,3}$, $S \to S_{1,3}S_{4,1}$ | $S \to S_{2,2}S_{4,2}$, $S \to S_{2,3}A_{5,1}$ | | | | |
| 3 | $S \to S_{1,1}S_{2,2}$, $S \to S_{1,2}A_{3,1}$ | $S \to S_{2,2}S_{4,1}$ | | | | |
| 2 | $S \to S_{1,1}S_{2,1}$ | $S \to S_{2,1}A_{3,1}$ | | $S \to S_{4,1}A_{5,1}$ | | |
| 1 | $S \to a$ | $S \to a$ | $A \to c$ | $S \to a$ | $A \to c$ | |
| | 1 | 2 | 3 | 4 | 5 | $i$ |

7 pts

**Question 7 (BA LL(1) (10 pts))**

1. *Assume that we want to do LL(1) parsing, using the precompiled* $First$ *and* $Follow$ *sets for a given grammar.*

   *Explain in your own words how these sets constrain the possible predicts that we can make in a situation where our top-most stack symbol is a non-terminal and we use the next input symbol as lookahead.*

2. *Consider the CFG* $G$ *with* $N = \{S, X\}, T = \{a, b, d\}$*, start symbol* $S$ *and productions*

$$S \to Xb, X \to aXd \mid \varepsilon$$

   (a) *Compute* $First(Xb)$.

   (b) *Compute* $Follow(X)$.

   (c) *Is this grammar LL(1)?*
   *Hint: you only have to check whether for the non-terminal* $X$*, LL(1) allows to decide deterministically which of the two* $X$*-productions to predict.*

Solution:

1. For a non-terminal $A$ on the stack, we predict a production $A \to \alpha$ from the grammar only if one of the following two cases holds:

   (a) Either the next input symbol is in the First set of $\alpha$

   (b) or $\varepsilon$ is in the First set of $\alpha$ and the next input symbol is in the Follow set of $\alpha$.

   4 points

2. (a) $First(Xb) = \{a, b\}$. 1 pt

   (b) $Follow(X) = \{b, d\}$. 2 pt1

   (c) The grammar is LL(1) since $First(aXd) = \{a\}$ and $Follow(X) = \{b, d\}$ are disjoint.

   3 pts

**Question 7 (MA LL(1) (10 pts))**

1. *Assume that we want to do LL(1) parsing, using the precompiled First and Follow sets for a given grammar.*

   *The orginal top-down predict deduction rule is*

$$\textit{Predict: } \frac{[A\alpha, i]}{[\gamma\alpha, i]} \quad A \to \gamma \in P$$

   *Give the modified predict deduction rule that restricts possible predicts by taking the next input symbol into consideration the way it is done in LL(1) parsing.*

2. *Consider the CFG G with $N = \{S, X\}, T = \{a, b, d\}$, start symbol $S$ and productions*

$$S \to Xb \,|\, dX, X \to aXd \,|\, \varepsilon$$

   (a) *Compute $First(Xb)$.*

   (b) *Compute $Follow(X)$.*

   (c) *Is this grammar LL(1)?*

Solution:

1. Predict: $\frac{[A\alpha, i]}{[\gamma\alpha, i]} \quad A \to \gamma \in P, w_{i+1} \in First(\alpha)$ or $\varepsilon \in First(\alpha)$ and $w_{i+1} \in Follow(\alpha)$

4 points

2. (a) $First(Xb) = \{a, b\}$. 1 pt

   (b) $Follow(X) = \{b, d, \$\}$. 2 pts

   (c) The grammar is LL(1) since

   - $First(Xb) = \{a, b\}$ and $First(dX) = \{d\}$ are disjoint, and
   - $First(aXd) = \{a\}$ and $Follow(X) = \{b, d, \$\}$ are disjoint.

3 pts