Parsing
Mid term exam, 28.11.2017

Laura Kallmeyer

Winter 2017, 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.

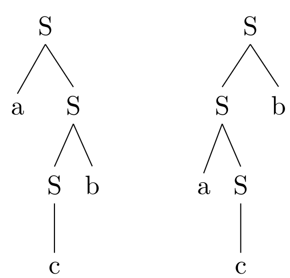
Question 1 (BA CFG, 9 pts)

Consider the CFG G with non-terminals $N = \{S\}$, terminals $T = \{a, b, c\}$, start symbol S and productions $S \to aS \,|\, Sb \,|\, c$.

1. *Is G in Chomsky Normal Form? Explain your answer.*
2. *Is G left-recursive? Explain your answer.*
3. *Give the two parse trees for acb that one obtains with G.*
4. *What is the language $L(G)$ generated by this grammar?*

Solution:

1. No ... 1 pt
2. Yes, since $S \Rightarrow Sb$ because of the production $S \to Sb$. 2 pts
3. 2 pts



4. The language can be characterized by a regular expression: a^*cb^* 4 pts

Question 1 (MA CFG, 9 pts)

Consider the CFG G with non-terminals $N = \{S, T\}$, terminals $T = \{a, b, c\}$, start symbol S and productions $S \to aTb \,|\, Tb, T \to S \,|\, c$.

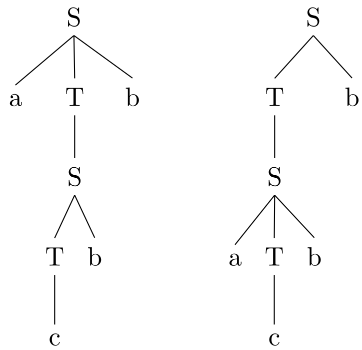
1. *Is G in Chomsky Normal Form? Explain your answer.*
2. *Is G left-recursive? Explain your answer.*
3. *Give the two parse trees for acbb that one obtains with G.*
4. *What is the language $L(G)$ generated by this grammar?*

Solution:

1. No, ... 1 pt

2. Yes, because $S \Rightarrow Tb \Rightarrow Sb$. 2 pts

3. 2 pts



4. $\{a^n cb^m \mid m \geq n \geq 0, m \geq 1\}$ 4 pts

Question 2 (PDA, 7 pts)

Give a PDA M that recognizes the language $\{a^n bc^m \mid n \geq m \geq 0\}$ when accepting with the empty stack, i.e., $L = N(M)$.

Hint: The stack has to keep track of how many as one sees before the b. When coming to the cs, for each a, we have at most one corresponding c.

Solution:

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

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

$\delta(q_1, c, A) = \{\langle q_1, \varepsilon \rangle\}$ $\quad \delta(q_1, \varepsilon, A) = \{\langle q_1, \varepsilon \rangle\}$ 7 pts

$\delta(q_1, \varepsilon, S) = \{\langle q_1, \varepsilon \rangle\}$

Question 3 (Top-Down parsing (8 pts)) *Consider the CFG G with $N = \{S, X\}, T = \{a\}$, start symbol S and productions*

$$S \to aSa \mid X, X \to aX \mid a$$

and the input $w = aaa$.

Give all pairs of prediction stack, remaining input and analysis stack that arise in a top-down parsing for this input. (The productions are assumed to be numbered in the order above, i.e., $S \to aSa$ is the 1st S-production etc.) List them in a table with a unique number for each pair and indicate from which other pair 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*	*analysis stack*	*obtained from*
1.	S	*aaa*	–	*axiom*
2.	*aSa*	*aaa*	S_1	*from 1.*
3.	X	*aaa*	S_2	*from 1.*
4.	...	...		

Solution:

id	stack	rem. input	analysis stack	obtained from
1.	S	aaa	–	axiom
2.	aSa	aaa	S_1	from 1.
3.	X	aaa	S_2	from 1.
4.	Sa	aa	S_1a	from 2.
5.	aX	aaa	S_2X_1	from 3.
6.	a	aaa	S_2X_2	from 3.
7.	Xa	aa	S_1S_2	from 4.
8.	X	aa	S_2X_1a	from 5.
9.	–	aa	S_2X_2a	from 6.
10.	aa	aa	$S_1S_2X_2$	from 7.
11.	aX	aa	$S_2X_1aX_1$	from 8.
12.	a	aa	$S_2X_1aX_2$	from 8.
13.	a	a	$S_1S_2X_2a$	from 10.
14.	X	a	$S_2X_1aX_1a$	from 11.
15.	–	a	$S_2X_1aX_2a$	from 12.
16.	–	–	$S_1S_2X_2aa$	from 13.
17.	a	a	$S_2X_1aX_1aX_2$	from 14.
18.	–	–	$S_2X_1aX_1aX_2a$	from 17.

(8 pts)

Question 4 (Top-down parsing with deduction rules (5 pts))

Consider the deduction-based specification of top-down parsing with additional constraints that the stack must not be longer than the remaining input (for ε-free grammars) and that for predicts, the next input symbol must be in the First set of the righthand side of the predicted production:

Axiom: $\frac{}{[S,0]}$ *S start symbol* *Scan:* $\frac{[a\alpha, i]}{[\alpha, i+1]}\ w_{i+1} = a$

Predict: $\frac{[A\alpha, i]}{[\gamma\alpha, i]}\ A \to \gamma \in P, |\gamma\alpha| \leq n - i, w_{i+1} \in First(\gamma)$

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

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

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

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*
$[S,0]$	$[S,0]$
...	...

Solution:

Input *bb*

new chart items	agenda items
$[S,0]$	$[S,0]$
$[SS,0]$, $[SA,0]$, $[b,0]$	$[SS,0]$, $[SA,0]$, $[b,0]$
$[bS,0]$	$[SA,0]$, $[b,0]$, $[bS,0]$,
$[bA,0]$	$[b,0]$, $[bS,0]$, $[bA,0]$
$[\varepsilon,1]$	$[bS,0]$, $[bA,0]$, $[\varepsilon,1]$
$[S,1]$	$[bA,0]$, $[\varepsilon,1]$, $[S,1]$
$[A,1]$	$[\varepsilon,1]$, $[S,1]$, $[A,1]$
	$[S,1]$, $[A,1]$
$[b,1]$	$[A,1]$, $[b,1]$
	$[b,1]$
$[\varepsilon,2]$	

(5 pts)

Question 5 (CYK-Parsing with dotted productions (7 pts)) *Consider the CYK parser with dotted productions that allows also for ε-productions.*[1]

We assume in addition that an item $[A \to \alpha \bullet \beta, i, j]$ is only generated if either $w_{j+1} \in First(\beta)$ or $\varepsilon \in First(\beta)$.

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

$$S \to aSa \,|\, aS \,|\, bSb \,|\, \varepsilon$$

and the input $w = abb$.

Give the chart one obtains when parsing w with this version of bottom-up parsing with dotted productions.

(Note that the chart is indexed with the start and end positions of the spans, i and j.)

Solution:

j					
3	$S \to aS\bullet$	$S \to bSb\bullet$		$S \to \bullet$	
2		$S \to b \bullet Sb$, $S \to bS \bullet b$	$S \to \bullet bSb$, $S \to \bullet$		
1	$S \to a \bullet Sa$, $S \to a \bullet S$, $S \to aS\bullet$	$S \to \bullet bSb$, $S \to \bullet$			
0	$S \to \bullet aSa$, $S \to \bullet aS$, $S \to \bullet$				
	0	1	2	3	i

(7pts)

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

$$S \to AS \,|\, SA \,|\, AA, A \to a$$

and the input $w = aaa$.

Give the chart one obtains when parsing w with the CNF version of CYK that put span-annotated productions into the chart.

(Be careful: this time, we are using i and l as chart indices.)

Solution:

[1] Goal items: $[S \to \alpha\bullet, 0, n]$ for all S-productions $S \to \alpha$.

Predict (axioms): $\dfrac{}{[A \to \bullet\alpha, i, i]}$ $A \to \alpha \in P, i \in [0..n]$

Scan: $\dfrac{[A \to \alpha \bullet a\beta, i, j]}{[A \to \alpha a \bullet \beta, i, j+1]}$ $w_{j+1} = a$ Complete: $\dfrac{[A \to \alpha \bullet B\beta, i, j][B \to \gamma\bullet, j, k]}{[A \to \alpha B \bullet \beta, i, k]}$

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

(4pts)

Question 6 (MA CYK-Parsing (4 pts)) *Show the soundness of the CYK algorithm as given in the deduction-based form for CNF-grammars on slide 12*[2]*. In other words, show that, given a CFG G in CNF and an input word w that is parsed, the following holds: If an item $[A, i, l]$ can be deduced, then $A \stackrel{*}{\Rightarrow} w_i \dots w_{i+l-1}$.*

Hint: Show this by induction over the applied deduction rules.

Solution:

- Axioms (scan): $[A, i, 1]$ is an axiom if $A \to w_i \in P$, consequently $A \Rightarrow w_i$.
- Complete: Assume that we have items $[B, i, l_1], [C, i + l_1, l_2]$ and our induction claim holds for these items, i.e., $B \stackrel{*}{\Rightarrow} w_i \dots w_{i+l_1-1}$ and $C \stackrel{*}{\Rightarrow} w_{i+l_1} \dots w_{i+l_1+l_2-1}$. Furthermore, assume that $A \to BC \in P$. Then we can deduce $[C, i, l_1 + l_2]$ and it holds that $A \Rightarrow BC \stackrel{*}{\Rightarrow} w_i \dots w_{i+l_1-1}C \stackrel{*}{\Rightarrow} w_i \dots w_{i+l_1-1}w_{i+l_1} \dots w_{i+l_1+l_2-1} = w_i \dots w_{i+l_1+l_2-1}$.

Since all items generated by the parser are either an axiom or obtained from the axioms by a sequence of complete steps, every item necessarily satisfies our soundness claim.

Question 7 (Shift-reduce (4 pts)) *Consider again the CFG G with $N = \{S\}, T = \{a\}$, start symbol S and productions*

$$S \to SS \mid a$$

and the input $w = aaa$.

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.	ε	*aaa*	*axiom*
2.	*a*	*aa*	*shift from 1.*
3.	*S*	*aa*	*reduce from 2. with $S \to a$*
	...	...	

Solution:

[2]Goal item: $[S, 1, n]$

Scan: $\dfrac{}{[A, i, 1]}\ A \to w_i \in P$ Complete: $\dfrac{[B, i, l_1], [C, i + l_1, l_2]}{[A, i, l_1 + l_2]}\ A \to B\ C \in P$

id	stack	rem. input	operation
1.	ε	aaa	axiom
2.	a	aa	shift from 1.
3.	S	aa	reduce from 2. with $S \to a$
4.	Sa	a	shift from 3.
5.	SS	a	reduce from 4. with $S \to a$
6.	S	a	reduce from 5. with $S \to SS$
7.	SSa	–	shift from 5.
8.	Sa	–	shift from 6
9.	SSS	–	reduce from 7. with $S \to a$
10.	SS	–	reduce from 7. with $S \to a$ (or from 9. with $S \to SS$)
11.	S	–	reduce from 10. with $S \to SS$

Question 8 (LL(1) (6 pts))

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

$$S \to aXA \,|\, aAb, X \to c, A \to aaA \,|\, Ac$$

1. *Compute $First(A)$.*
2. *Compute $Follow(A)$.*
3. *Is this grammar LL(1)? Explain your answer.*

Solution:

1. $First(A) = \{a\}$ (1pt)
2. $Follow(A) = \{b, c, \$\}$ (3pts)
3. No, since $First(aXA) = First(aAb) = \{a\}$ concerning the two S-productions and also, concerning the A-productions, $First(aaA) = \{a\}$ and $First(Ac) = First(A) = \{a\}$. (2pts)