

Automatische Syntaxanalyse (Parsing)

Abschlussklausur

28.01.2013

Laura Kallmeyer

WS 2012/13, Heinrich-Heine-Universität Düsseldorf

Dauer der Abschlussklausur: 90 Minuten.

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

Aufgabe 1 (Left-Corner Parsing, 8 Punkte) Betrachten Sie die CFG $\langle N, T, P, S \rangle$ mit $N = \{S\}$, $T = \{a, c\}$, Startsymbol S , Produktionen $S \rightarrow aSc \mid aS \mid c$.

1. Welche Sprache wird von dieser Grammatik generiert?
2. Geben Sie die Tripel von Stacks Γ_{compl} , Γ_{td} und Γ_{lhs} an, die sich bei einem Left-Corner-Parsing (nur Recognition) für die Eingabe *acc* ergeben. Zeigen Sie dabei, welches Tripel aus welchem anderen entstanden ist und mit welcher Operation jeweils:

	Γ_{compl}	Γ_{td}	Γ_{lhs}	Operation
1.	<i>acc</i>	S	ε	
2.	<i>cc</i>	$ScSS$	S	<i>reduce(1)</i>
3.	<i>cc</i>	SSS	S	<i>reduce(1)</i>
				\dots

Lösung:

1. $\{a^n c^m c \mid 0 \leq m \leq n\}$

	Γ_{compl}	Γ_{td}	Γ_{lhs}	
1.	<i>acc</i>	S	ε	
2.	<i>cc</i>	$ScSS$	S	<i>reduce(1)</i>
3.	<i>cc</i>	SSS	S	<i>reduce(1)</i>
4.	<i>c</i>	$\$ScSS$	SS	<i>reduce(2)</i>
5.	<i>c</i>	$\$SSS$	SS	<i>reduce(3)</i>
6.	Sc	$ScSS$	S	<i>move(4)</i>
7.	Sc	SSS	S	<i>move(5)</i>
8.	<i>c</i>	cSS	S	<i>remove(6)</i>
2. 9.	<i>c</i>	$\$S$	S	<i>remove(7)</i>
10.	ε	$\$S$	S	<i>remove(8)</i>
11.	ε	$\$cSS$	SS	<i>reduce(8)</i>
12.	Sc	S		<i>move(9)</i>
13.	S	S		<i>move(10)</i>
14.	S	cSS	S	<i>move(11)</i>
15.	<i>c</i>	ε		<i>remove(12)</i>
16.	ε	ε	ε	<i>remove(13)</i>
17.	ε	$\$$	S	<i>reduce(15)</i>
18.	S	ε	ε	<i>move(17)</i>

3. Der Parser hat eine Konfiguration mit drei leeren Stacks (16) erreicht, folglich ist das Wort in der Sprache.

Aufgabe 2 (Earley Parsing, 13 Punkte)

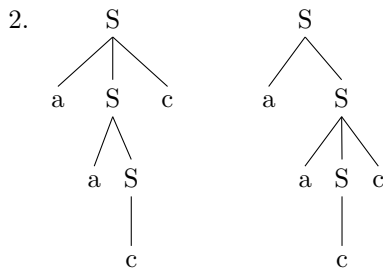
Betrachten Sie erneut die CFG aus der vorhergehenden Aufgabe.

- Geben Sie die Chart an, die sich bei einem Earley-Parsing mit Backpointer und mit den Lookahead Bedingungen auf Predict und Complete¹ mit dieser Grammatik für die Eingabe aacc ergibt.
- Welche Parsbäume ergeben sich für aacc?

Lösung:

- Chart (hier ohne Backpointer, die mssten noch dazu kommen):

j					
	$S \rightarrow aS\bullet$				
4	$S \rightarrow aSc\bullet$	$S \rightarrow aSc\bullet$			
	$S \rightarrow aS\bullet$	$S \rightarrow aS\bullet$			
3	$S \rightarrow aS\bullet c$	$S \rightarrow aS\bullet c$	$S \rightarrow c\bullet$		
		$S \rightarrow a\bullet S$			
2		$S \rightarrow a\bullet Sc$	$S \rightarrow \bullet c$		
	$S \rightarrow a\bullet S$	$S \rightarrow \bullet aS$			
1	$S \rightarrow a\bullet Sc$	$S \rightarrow \bullet aSc$			
	$S \rightarrow \bullet c$				
0	$S \rightarrow \bullet aS$				
	$S \rightarrow \bullet aSc$				
	0	1	2	3	4 i



Aufgabe 3 (LR Parsing, 14 Punkte)

Betrachten Sie die CFG $\langle N, T, P, S \rangle$ mit

$N = \{S\}$, $T = \{a, c\}$, Startsymbol S , Produktionen 1. $S \rightarrow aSc$, 2. $S \rightarrow aS$, 3. $S \rightarrow \epsilon$.

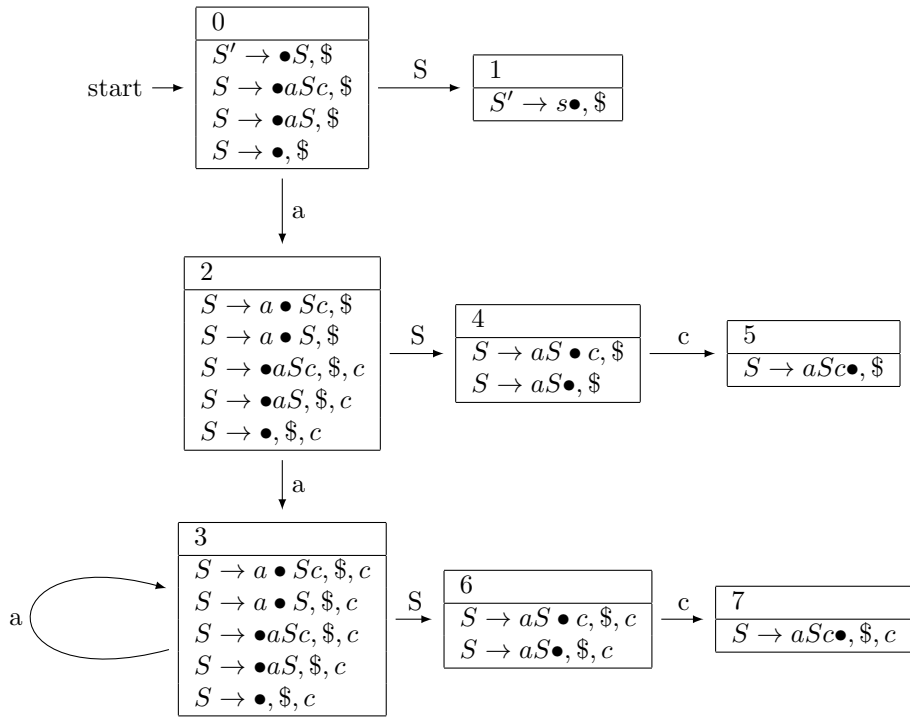
- Geben Sie Zustände und Übergänge an, die sich mit einer kanonischen LR(1)-Konstruktion (canonical LR) ergeben und lesen Sie die Parsingtable ab.
- Ist die Grammatik LR(1)? Begründen Sie Ihre Antwort.

Lösung:

- Automat:

$$\frac{\text{¹Predict with lookahead: } [A \rightarrow \alpha \bullet B\beta, i, j]}{[B \rightarrow \bullet \gamma, j, j]} \quad B \rightarrow \gamma \in P, w_{j+1} \in \text{First}(\gamma) \text{ or } \epsilon \in \text{First}(\gamma)$$

$$\frac{\text{Complete with lookahead: } [A \rightarrow \alpha \bullet B\beta, i, j], [B \rightarrow \gamma \bullet, j, k]}{[A \rightarrow \alpha B \bullet \beta, i, k]} \quad \begin{array}{l} w_{k+1} \in \text{First}(\beta) \\ \text{or } \epsilon \in \text{First}(\beta) \text{ and } w_{k+1} \in \text{Follow}(A) \end{array}$$



	a	c	\$	S
0	s2		r3	1
1			acc	
2	s3	r3	r3	4
3	s3	r3	r3	6
4		s5	r2	
5			r1	
6		s7, r2	r2	
7		r1	r1	

2. Nein, da im Feld 6, c zwei Einträge stehen.

Aufgabe 4 (Tomita's LR Parser, 7 Punkte)

Betrachten Sie die CFG $\langle N, T, P, S \rangle$ mit

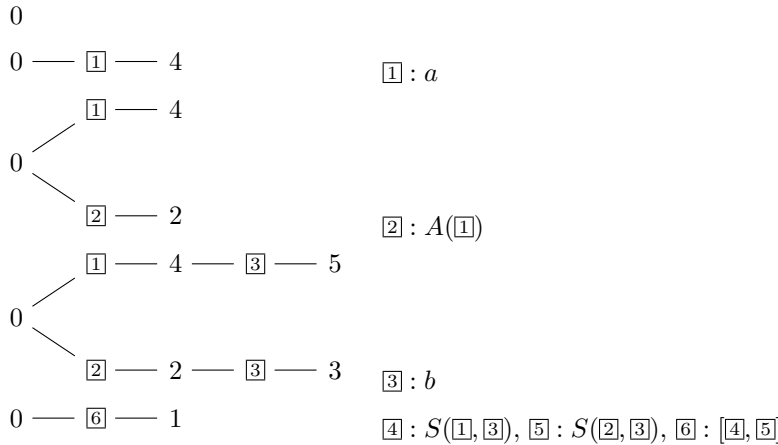
$N = \{S, A\}$, $T = \{a, b\}$, Startsymbol S , Produktionen $1.S \rightarrow Ab$, $2.S \rightarrow ab$, $3.A \rightarrow aA$, $4.A \rightarrow a$.

Bei der kanonischen LR(1) Konstruktion ergibt sich für diese Grammatik die folgende Parsingtabelle:

	a	b	\$	A	S
0	s4			2	1
1			acc		
2		s3			
3			r1		
4	s6	s5, r4		7	
5			r2		
6	s6	r4		7	
7		r3			

Geben Sie für die Eingabe ab die Abfolge der Stackgraphen mit dem zugehörigen Parswald an, die sich bei Tomitas Parser für diese Eingabe mit der angegebenen Grammatik und Tabelle ergibt.

Lösung:



Aufgabe 5 (Weighted Deductive Parsing, 8 Punkte)

Betrachten Sie das Earley Parsing Deduktionssystem, versehen mit zusätzlichen Gewichten:

Axioms: $\frac{}{\langle |\log(p)|, |\log(p)| \rangle : [S \rightarrow \bullet \alpha, 0, 0]} \quad p : S \rightarrow \alpha \in P, w_1 \in First(\alpha)$

Scan: $\frac{\langle x, y \rangle : [A \rightarrow \alpha \bullet a \beta, i, j]}{\langle x, y \rangle : [A \rightarrow \alpha a \bullet \beta, i, j + 1]} \quad w_{j+1} = a$

Predict: $\frac{\langle x, y \rangle : [A \rightarrow \alpha \bullet B \beta, i, j]}{\langle x + |\log(p)|, |\log(p)| \rangle : [B \rightarrow \bullet \gamma, j, j]} \quad p : B \rightarrow \gamma \in P, w_{j+1} \in First(\gamma)$

Complete: $\frac{\langle x_1, y_1 \rangle : [A \rightarrow \alpha \bullet B \beta, i, j], \langle x_2, y_2 \rangle : [B \rightarrow \gamma \bullet, j, k]}{\langle x_1 + y_2, y_1 + y_2 \rangle : [A \rightarrow \alpha B \bullet \beta, i, k]}$

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

Die Gewichte sind wie folgt geordnet:

$\langle x_1, y_1 \rangle < \langle x_2, y_2 \rangle$ genau dann wenn entweder $x_1 < x_2$ oder $x_1 = x_2$ und $y_1 < y_2$.

Betrachten Sie die PCFG $\langle N, T, P, S \rangle$ mit

$N = \{S\}$, $T = \{a, b, c\}$, Startsymbol S , Produktionen

$$0.6 : S \rightarrow aSb, \quad 0.1 : S \rightarrow aSc, \quad 0.1 : S \rightarrow a, \quad 0.2 : S \rightarrow c$$

Es gilt (gerundet) $|\log(0.6)| = 0.22$, $|\log(0.1)| = 1$ und $|\log(0.2)| = 0.7$.

Zeigen Sie in Form einer Tabelle, wie sich Chart und Agenda schrittweise entwickeln, wenn mit dieser Grammatik und dem angegebenen gewichteten Earley Deduktionssystem die Eingabe $w = acb$ geparkt wird. (Ohne Backpointer, also nur Recognition.)

Beachten Sie dabei die Bedingungen bzgl. der First Menge dessen, was rechts vom Dot steht und des nächsten Eingabesymbols.

Es genügt im Prinzip, die schrittweise Entwicklung der Agenda zu zeigen, da sich die Chart unmittelbar hieraus ergibt. (Die Chart enthält alles, was im Laufe des Parsens aus der Agenda entfernt wird.)

Lösung:

Chart	Agenda
	$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$	$\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$	$\langle 0.92, 0.7 \rangle : [S \rightarrow \bullet c, 1, 1]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow \bullet c, 1, 1]$	$\langle 0.92, 0.7 \rangle : [S \rightarrow c \bullet, 1, 2]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow \bullet c, 1, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow c \bullet, 1, 2]$	$\langle 0.92, 0.92 \rangle : [S \rightarrow aS \bullet b, 0, 2]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow \bullet c, 1, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow c \bullet, 1, 2]$, $\langle 0.92, 0.92 \rangle : [S \rightarrow aS \bullet b, 0, 2]$	$\langle 0.92, 0.92 \rangle : [S \rightarrow aSb \bullet, 0, 3]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$
$\langle 0.22, 0.22 \rangle : [S \rightarrow \bullet aSb, 0, 0]$, $\langle 0.22, 0.22 \rangle : [S \rightarrow a \bullet Sb, 0, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow \bullet c, 1, 1]$, $\langle 0.92, 0.7 \rangle : [S \rightarrow c \bullet, 1, 2]$, $\langle 0.92, 0.92 \rangle : [S \rightarrow aS \bullet b, 0, 2]$, $\langle 0.92, 0.92 \rangle : [S \rightarrow aSb \bullet, 0, 3]$	$\langle 1, 1 \rangle : [S \rightarrow \bullet aSc, 0, 0]$, $\langle 1, 1 \rangle : [S \rightarrow \bullet a, 0, 0]$