Einführung in die Computerlinguistik
Abschlussklausur

Laura Kallmeyer, Simon Petitjean

Sommersemester 2016, Heinrich-Heine-Universität Düsseldorf

18.07.2016

Erlaubte Hilfsmittel: Eine Din-A4 Seite mit Notizen. Kein Taschenrechner.

Aufgabe 1 (8 Pkte) *Consider the following CFG:*
$G = \langle \{S, A, B\}, \{a, b\}, \{S \to aBA \mid aBAS, A \to a \mid aS, B \to aBbA \mid b\}, S \rangle$

1. *Is this grammar in Greibach normal form? Justify your answer.*
2. *Give a left derivation and a right derivation for $w = aabbaa$ (there is only one for each).*
3. *Give all the possible parse trees for $w = abaaba$.*
4. *Is G an ambiguous grammar?*

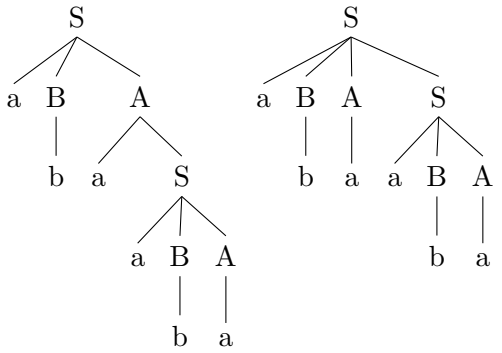
Solution:

1. No, because the rule $B \to aBbA$ has a second terminal symbol (b), while the sequence after a should only be composed of non terminal symbols.

   1 Pkt

2. Left derivation: $S \Rightarrow aBA \Rightarrow aaBbAA \Rightarrow aabbAA \Rightarrow aabbaA \Rightarrow aabbaa$
   Right derivation: $S \Rightarrow aBA \Rightarrow aBa \Rightarrow aaBbAa \Rightarrow aaBbaa \Rightarrow aabbaa$ 2 Pkte

3. 4 Pkte



4. Yes, because the previous input has more than one derivation tree. 1 Pkt

Aufgabe 2 (8 Pkte) *Consider the following feature structures, expressed as attibute-value matrices:*

$$S_1 = \begin{bmatrix} \textit{sentence} & \\ \text{SUBJ} & \begin{bmatrix} \textit{nominal} & \\ \text{CASE} & \boxed{1} \\ \text{AGR} & \boxed{4} \end{bmatrix} \\ \text{PRED} & \begin{bmatrix} \textit{verbal} & \\ \text{SUBJCASE} & \boxed{1} \\ \text{ASSIGNCASE} & \boxed{3} \\ \text{AGR} & \boxed{4} \\ \text{SCOMP} & \begin{bmatrix} \textit{sentence} & \\ \text{SUBJ} & \begin{bmatrix} \textit{nominal} & \\ \text{CASE} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$S_2 = \begin{bmatrix} \textit{sentence} & \\ \text{PRED} & \begin{bmatrix} \textit{ecm} & \\ \text{SUBJCASE} & \text{nom} \\ \text{ASSIGNCASE} & \text{acc} \\ \text{PHON} & \text{expects} \\ \text{AGR} & \begin{bmatrix} \textit{agreement} & \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

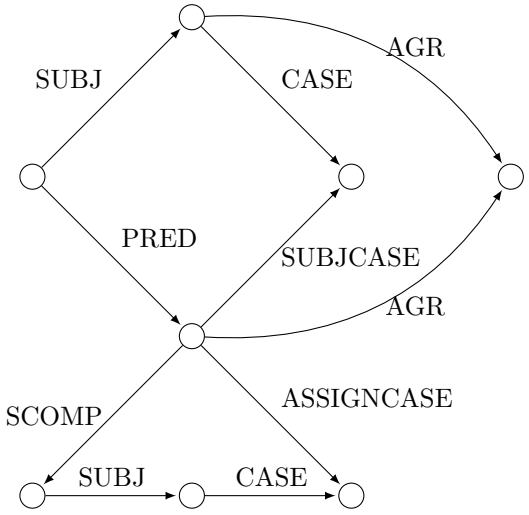
$$S_3 = \begin{bmatrix} \textit{sentence} & \\ \text{SUBJ} & \begin{bmatrix} \textit{pronoun} & \\ \text{CASE} & \text{acc} \\ \text{PHON} & \text{him} \\ \text{AGR} & \begin{bmatrix} \textit{agreement} & \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Types: ecm *is a subtype of* verbal *and* pronoun *a subtype of* nominal.

1. *Give the equivalent graph for* S_1.
2. *Calculate* $S_1 \sqcup S_2$, $S_1 \sqcup S_3$ *and* $S_1 \sqcup S_2 \sqcup S_3$. *If the unification is not possible (result* $\perp$*), explain why the unification fails.*

Solution:

1.



2. $S_1 \sqcup S_2 =$

$$\begin{bmatrix} \textit{sentence} \\ \text{SUBJ} & \begin{bmatrix} \textit{nominal} \\ \text{CASE} & \boxed{1}\text{nom} \\ \text{AGR} & \boxed{4}\begin{bmatrix} \textit{agreement} \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix} \\ \text{PRED} & \begin{bmatrix} \textit{ecm} \\ \text{SUBJCASE} & \boxed{1}\text{nom} \\ \text{ASSIGNCASE} & \boxed{3}\text{acc} \\ \text{PHON} & \text{expects} \\ \text{AGR} & \boxed{4}\begin{bmatrix} \textit{agreement} \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SCOMP} & \begin{bmatrix} \textit{sentence} \\ \text{SUBJ} & \begin{bmatrix} \textit{nominal} \\ \text{CASE} & \boxed{3}\text{acc} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$S_1 \sqcup S_3 =$

$$\begin{bmatrix} \textit{sentence} \\ \text{SUBJ} & \begin{bmatrix} \textit{pronoun} \\ \text{CASE} & \boxed{1}\text{acc} \\ \text{PHON} & \text{him} \\ \text{AGR} & \boxed{4}\begin{bmatrix} \textit{agreement} \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix} \\ \text{PRED} & \begin{bmatrix} \textit{verbal} \\ \text{SUBJCASE} & \boxed{1}\text{acc} \\ \text{ASSIGNCASE} & \boxed{3} \\ \text{AGR} & \boxed{4}\begin{bmatrix} \textit{agreement} \\ \text{NUM} & \text{sg} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SCOMP} & \begin{bmatrix} \textit{sentence} \\ \text{SUBJ} & \begin{bmatrix} \textit{nominal} \\ \text{CASE} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$S_1 \sqcup S_2 \sqcup S_3 = \bot$ because of the different value for CASE under SUBJ.

Aufgabe 3 (6 Pkte) *Consider the following CFG G : $N = \{S, A, B\}$, $T = \{a, b\}$, with start symbol S.*

Productions:

$S \to A \quad S \to B \quad A \to bB \quad A \to aAa \quad B \to b \quad B \to bB$

Give the trace obtained by a Top-Down-Parsing of the input $w = abba$, e.g. all the successive couples of stack and remaining input. Indicate for each step by which operation and from which other couple it was created. Please do not use the length check proposed in the homework (you have to continue even if the stack is longer than the input).

Solution:

	Remaining input	Stack	
1.	abba	S	
2.	abba	A	predict(1)
3.	abba	B	predict(1)
4.	abba	bB	predict(2)
5.	abba	aAa	predict(2)
6.	abba	b	predict(3)
7.	abba	bB	predict(3)
8.	bba	Aa	scan(5)
9.	bba	bBa	predict(8)
10.	bba	aAaa	predict(8)
11.	ba	Ba	scan(9)
12.	ba	ba	predict(11)
13.	ba	bBa	predict(11)
14.	a	a	scan(12)
15.	a	Ba	scan(13)
16.	a	ba	predict(15)
17.	a	bBa	predict(15)
18.	ε	ε	scan(14)

Aufgabe 4 (6 Pkte) *Consider the following CFG: $S \to aSd|aAd$, $A \to b|bc$*

1. *Give the trace of a shift-reduce parsing for the input $w = aabcdd$.*

   Give the trace as a table showing the index of the step, the remaining input, the stack, the operation (shift or reduce) used to obtain the configuration, and the index of the previous configuration (before application of the operation).

Index	*Remaining input*	*Stack*	*Operation*	*from*
1.	*aabcdd*	ϵ	–	–
2.	*abcdd*	*a*	*shift*	*1*

   During the parsing, when shift and reduce are both possible (shift/reduce conflict), start by trying the reduce possibility. If this choice does not lead to a successful parse, try the shift solution.

2. *How does the parser recognize that the word belongs to the language?*

Solution:

1.

Index	Remaining input	Stack	Operation	from	
1.	aabcdd	ϵ	–	–	
2.	abcdd	a	shift	1	
3.	bcdd	aa	shift	2	
4.	cdd	aab	shift	3	
5.	cdd	aaA	reduce	4	$\to$ fail
6.	dd	aabc	shift	4	
7.	dd	aaA	reduce	6	
8.	d	aaAd	shift	7	
9.	d	aS	reduce	8	
10.	ϵ	aSd	shift	9	
11.	ϵ	S	reduce	10	

2. Because the stack contains only the axiom S and the remaining input is empty.

Aufgabe 5 (6 Pkte) *Consider the following CFG:*
$G = \langle \{S, A, B, C, D, E\}, \{a, b, c\}, \{S \to aSD \,|\, aCA \,|\, aAc, A \to BD \,|\, B \,|\, aAB, B \to b, C \to abD, E \to aAc\}, S\rangle$

1. *Give an equivalent grammar without useless symbols. Please give the set P of productive symbols.*
2. *Which language is generated by this grammar?*

Solution:

1. a) Productive symbols: $\{a, b, c, B, A, S, E\}$
   New productions: $S \to aAc, A \to B \,|\, aAB, B \to b, E \to aAc$

   Reacheable symbols: $\{S, a, A, c, B, b\}$
   New productions $S \to aAc, A \to B \,|\, aAB, B \to b$

   b) $\{a^n b^n c \,|\, n > 0\}$

Aufgabe 6 (8 Pkte) *Consider the following CFG:*
$G = \langle \{S, A\}, \{a, b, c\}, \{S \to aSbA \,|\, abA, A \to abc\}, S\rangle$

Give an equivalent grammar in Chomsky Normal Form.

Solution:

1. Lexicalization:

   $S \to C_A S C_B A \,|\, C_A C_B A, A \to C_A C_B C_C, C_A \to a, C_B \to b, C_C \to c$

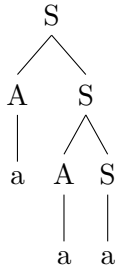
2. Binarization:

   $S \to C_A D_1 \,|\, C_A D_2, A \to C_A D_3, C_A \to a, C_B \to b, C_C \to c, D_2 \to C_B A, D_3 \to C_B C_C, D_1 \to S D_2$

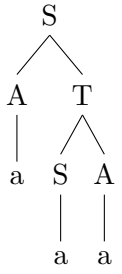
Aufgabe 7 (8 Pkte) *Consider the following PCFG:*
$G = \langle \{S, A, T\}, \{a\}, \{0.6 : S \to AS, 0.3 : S \to AT, 0.1 : S \to a, 1 : A \to a, 1 : T \to SA\}, S\rangle$

1. *Give the two parse trees for aaa with the corresponding probabilities.*
2. *Give the inside chart for aaa.*

Solution:

1. 

, probability $36 \cdot 10^{-3}$



, probability $3 \cdot 10^{-2}$

2. Inside chart:

j				
3	$\langle S, 0.066\rangle$, $\langle T, 0.06\rangle$	$\langle S, 0.06\rangle$, $\langle T, 0.1\rangle$	$\langle A, 1\rangle$, $\langle S, 0.1\rangle$	
2	$\langle S, 0.06\rangle$, $\langle T, 0.1\rangle$	$\langle A, 1\rangle$, $\langle S, 0.1\rangle$		
1	$\langle A, 1\rangle$, $\langle S, 0.1\rangle$			
	1	2	3	i