Parsing Beyond CFG
Homework 8: Linear Context-free Rewriting Systems (LCFRS) Parsing

Laura Kallmeyer

Question 1 *Consider the LCFRS/simple RCG with the following rules:*

$S(XYZU) \rightarrow A(X,Z)B(U,Y)$	$S(XYZ) \rightarrow A(X,Z)C(Y)$
$A(aX,aZ) \rightarrow A(X,Z)$	$A(\varepsilon,c) \rightarrow \varepsilon$
$B(Xb,Yb) \rightarrow B(X,Y)$	$B(\varepsilon,c) \rightarrow \varepsilon$
$C(aXY) \rightarrow D(X)C(Y)$	$D(d) \rightarrow \varepsilon$

1. *Perform the following transformations on this simple RCG while obtaining always weakly equivalent simple RCGs:*
   (a) *Transform the grammar into an ordered simple RCG.*
   (b) *Remove useless rules.*
   (c) *Remove ε-rules.*
2. *What is the string language generated by this grammar?*

Solution:

1. Simplifying the grammar:
   (a) Transform the grammar into an ordered simple RCG. (If the superscript is the identity, we omit it.)
   The only problematic rule is $S(XYZU) \rightarrow A(X,Z)B(U,Y)$. It transforms into $S(XYZU) \rightarrow A(X,Z)B^{\langle 2,1\rangle}(Y,U)$.
   Add $B^{\langle 2,1\rangle}(Yb,Xb) \rightarrow B(X,Y)$ and $B^{\langle 2,1\rangle}(c,\varepsilon) \rightarrow \varepsilon$.
   Then, $B^{\langle 2,1\rangle}(Yb,Xb) \rightarrow B(X,Y)$ transforms into $B^{\langle 2,1\rangle}(Yb,Xb) \rightarrow B^{\langle 2,1\rangle}(Y,X)$.
   In the following, for reasons of readability, we replace $B^{\langle 2,1\rangle}$ with a new symbol E.
   Result:

$S(XYZU) \rightarrow A(X,Z)E(Y,U)$	$S(XYZ) \rightarrow A(X,Z)C(Y)$
$A(aX,aZ) \rightarrow A(X,Z)$	$A(\varepsilon,c) \rightarrow \varepsilon$
$B(Xb,Yb) \rightarrow B(X,Y)$	$B(\varepsilon,c) \rightarrow \varepsilon$
$E(Yb,Xb) \rightarrow E(Y,X)$	$E(c,\varepsilon) \rightarrow \varepsilon$
$C(aXY) \rightarrow D(X)C(Y)$	$D(d) \rightarrow \varepsilon$

   (b) Remove useless rules.
   - $N_T = \{A,B,E,D,S\}$. Consequently, remove $S(XYZ) \rightarrow A(X,Z)C(Y)$ and $C(aXY) \rightarrow D(X)C(Y)$.
   - In the result, $N_S = \{S,A,E\}$. Consequently, remove also $D(d) \rightarrow \varepsilon$, $B(Xb,Yb) \rightarrow B(X,Y)$ and $B(\varepsilon,c) \rightarrow \varepsilon$.

   Result:

$S(XYZU) \rightarrow A(X,Z)E(Y,U)$	
$A(aX,aZ) \rightarrow A(X,Z)$	$A(\varepsilon,c) \rightarrow \varepsilon$
$E(Yb,Xb) \rightarrow E(Y,X)$	$E(c,\varepsilon) \rightarrow \varepsilon$

   (c) Remove ε-rules.
   $N_\varepsilon = \{A^{01},A^{11},E^{10},E^{11},S^{1}\}$.
   Resulting productions:

$$\begin{array}{ll} S^1(XYZU) \to A^{11}(X,Z)E^{11}(Y,U) & S^1(YZU) \to A^{01}(Z)E^{11}(Y,U) \\ S^1(XYZ) \to A^{11}(X,Z)E^{10}(Y) & S^1(YZ) \to A^{01}(Z)E^{10}(Y) \\ A^{11}(aX,aZ) \to A^{11}(X,Z) & A^{11}(a,aZ) \to A^{01}(Z) \\ A^{01}(c) \to \varepsilon & \\ E^{11}(Yb,Xb) \to E^{11}(Y,X) & E^{11}(Yb,b) \to E^{10}(Y) \\ E^{10}(c) \to \varepsilon & \end{array}$$

2. The string language generated by this grammar is

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

Question 2 *Consider the ordered simple RCG with the following rules:*

$$\begin{array}{ll} S(XYZ) \to A(X,Z)B(Y) & B(e) \to \varepsilon \\ A(aX,aY) \to A(X,Y) & A(c,c) \to \varepsilon \\ A(Xb,Yb) \to A(X,Y) & \end{array}$$

*Give **all items** that are deduced when parsing $w = cbecb$ with the **incremental Earley Parser** described in the course. List them in a table as follows:*

id	*rule*	*pos*	*bindings*	*operation*
1	$S(\bullet XYZ) \to A(X,Z)B(Y)$	*0*	*?,?,?*	*axiom*
2	$A(\bullet aX, aY) \to A(X,Y)$	*0*	*?,?,?,?*	*predict(1)*
3	$A(\bullet c, c) \to \varepsilon$	*0*	*?,?*	*predict(1)*
4	$A(\bullet Xb, Yb) \to A(X,Y)$	*0*	*?,?,?,?*	*predict(1)*
	...			

Solution:

id	rule	*pos*	bindings	operation
1	$S(\bullet XYZ) \to A(X,Z)B(Y)$	0	?,?,?	axiom
2	$A(\bullet aX, aY) \to A(X,Y)$	0	?,?,?,?	predict(1)
3	$A(\bullet c, c) \to \varepsilon$	0	?,?	predict(1)
4	$A(\bullet Xb, Yb) \to A(X,Y)$	0	?,?,?,?	predict(1)
5	$A(c\bullet, c) \to \varepsilon$	1	$\langle 0,1\rangle$,?	scan(3)
6	$S(X \bullet YZ) \to A(X,Z)B(Y)$	1	$\langle 0,1\rangle$,?,?	suspend(1,5)
7	$A(X \bullet b, Yb) \to A(X,Y)$	1	$\langle 0,1\rangle$,?,?,?	suspend(4,5)
8	$B(\bullet e) \to \varepsilon$	1	?	predict(6)
9	$A(Xb\bullet, Yb) \to A(X,Y)$	2	$\langle 0,1\rangle$,$\langle 1,2\rangle$,?,?	scan(7)
10	$S(X \bullet YZ) \to A(X,Z)B(Y)$	2	$\langle 0,2\rangle$,?,?	suspend(1,9)
11	$A(X \bullet b, Yb) \to A(X,Y)$	2	$\langle 0,2\rangle$,?,?,?	suspend(4,9)
12	$B(\bullet e) \to \varepsilon$	2	?	predict(10)
13	$B(e\bullet) \to \varepsilon$	3	$\langle 2,3\rangle$	scan(12)
14	$B(\langle 2,3\rangle)$			convert(13)
15	$S(XY \bullet Z) \to A(X,Z)B(Y)$	2	$\langle 0,2\rangle$,$\langle 2,3\rangle$,?	complete(10,14)
16	$A(Xb, \bullet Yb) \to A(X,Y)$	3	$\langle 0,1\rangle$,$\langle 1,2\rangle$,?,?	resume(9,15)
17	$A(c, \bullet c) \to \varepsilon$	3	$\langle 0,1\rangle$,?	resume(5,16)
18	$A(c, c\bullet) \to \varepsilon$	4	$\langle 0,1\rangle$,$\langle 3,4\rangle$	scan(17)
19	$A(\langle 0,1\rangle, \langle 3,4\rangle)$			convert(18)
20	$A(Xb, Y \bullet b) \to A(X,Y)$	4	$\langle 0,1\rangle$,$\langle 1,2\rangle$,$\langle 3,4\rangle$,?	complete(16,19)
21	$A(Xb, Yb\bullet) \to A(X,Y)$	5	$\langle 0,1\rangle$,$\langle 1,2\rangle$,$\langle 3,4\rangle$,$\langle 4,5\rangle$	scan(20)
22	$A(\langle 0,2\rangle, \langle 3,5\rangle)$			convert(21)
23	$S(XYZ\bullet) \to A(X,Z)B(Y)$	2	$\langle 0,2\rangle$,$\langle 2,3\rangle$,$\langle 3,5\rangle$	complete(15,22)
24	$S(\langle 0,5\rangle)$			convert(23)