Parsing Beyond CFG
Homework 10: LCFRS 3

Laura Kallmeyer

SS 2016, Heinrich-Heine-Universität Düsseldorf

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

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

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 $S(XYZU) \to A(X,Z)B(U,Y)$. It transforms into $S(XYZU) \to A(X,Z)B^{\langle 2,1\rangle}(Y,U)$.
   Add $B^{\langle 2,1\rangle}(Yb, Xb) \to B(X,Y)$ and $B^{\langle 2,1\rangle}(c, \varepsilon) \to \varepsilon$.
   Then, $B^{\langle 2,1\rangle}(Yb, Xb) \to B(X,Y)$ transforms into $B^{\langle 2,1\rangle}(Yb, Xb) \to B^{\langle 2,1\rangle}(Y,X)$.
   In the following, for reasons of readability, we replace $B^{\langle 2,1\rangle}$ with a new symbol E.
   Result:

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

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

   Result:

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

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

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

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 LCFRS/simple RCG with the following rules:*

$S(UXYZ) \to Comp(U)\, VP_{fin}(X,Z) N(Y)$ $\quad$ $N(Peter) \to \varepsilon$ $\quad$ $N(Maria) \to \varepsilon$ $\quad$ $N(Hans) \to \varepsilon$
$VP_{fin}(XY,ZU) \to VP_{inf}(X,Z) N(Y)\, V_{fin}(U)$ $\quad$ $Comp(dass) \to \varepsilon$ $\quad$ $V_{fin}(verspricht) \to \varepsilon$
$VP_{inf}(X,Y) \to N(X)\, V_{inf}(Y)$ $\quad$ $V_{inf}(zu\ treffen) \to \varepsilon$

Binarize this grammar with the deterministic left-to-right binarization that we have seen in the course.

Solution:

- Replace the rule $\mathrm{S}(UXYZ) \to \mathrm{Comp}(U)\mathrm{VP}_{fin}(X,Z)\mathrm{N}(Y)$ with the two new rules $\mathrm{S}(XY) \to \mathrm{Comp}(X)\mathrm{C}_1(Y)$ and $\mathrm{C}_1(XYZ) \to \mathrm{VP}_{fin}(X,Z)\mathrm{N}(Y)$;
- replace the rule $\mathrm{VP}_{fin}(XY,ZU) \to \mathrm{VP}_{inf}(X,Z)\mathrm{N}(Y)\mathrm{V}_{fin}(U)$ with $\mathrm{VP}_{fin}(XY,ZU) \to \mathrm{VP}_{inf}(X,Z)\mathrm{C}_2(Y,U)$ and $\mathrm{C}_2(Y,U) \to \mathrm{N}(Y)\mathrm{V}_{fin}(U)$.