Parsing Beyond CFG
Homework 7: Linear Context-free Rewriting Systems (LCFRS)

Laura Kallmeyer

Question 1

Consider the following LCFRS in sRCG format:

$$G = \langle \{A, B, C, S\}, \{a, b, c, d, e\}, \{X_1, X_2, X_3, Y_1, Y_2, Z_1, Z_2\}, P, S\rangle$$

where

$$\begin{aligned} P = \{ \quad & S(Z_1X_1Y_1X_2Y_2X_3Z_2) \rightarrow C(Z_1, Z_2)A(X_1, X_2, X_3)B(Y_1, Y_2), \\ & A(aX_1, bX_2, eX_3) \rightarrow A(X_1, X_2, X_3), \\ & B(dY_1, dY_2) \rightarrow B(Y_1, Y_2), \\ & A(a, b, e) \rightarrow \epsilon, \\ & B(d, d) \rightarrow \epsilon \\ & C(cc, cc) \rightarrow \varepsilon \quad \} \end{aligned}$$

1. What do the yields of A, B and C look like, given this grammar?
2. What is the string language generated by this LCFRS?
3. Give the same LCFRS in MCFG notation, i.e., with separate composition functions F that describe for each rule how to compute the yield of the righthand side non-terminal from the yields of the lefthand side non-terminals.

Solution:

1. yield(C) $= \{\langle cc, cc\rangle\}$;
   yield(B) $= \{\langle d^n, d^n\rangle \mid n \geq 1\}$;
   yield(A) $= \{\langle a^n, b^n, e^n\rangle \mid n \geq 1\}$
2. $L = \{cca^nd^mb^nd^me^ncc \mid n, m \geq 1\}$
3. $$G' = \langle \{A, B, C, S\}, \{a, b, c, d, e\}, F, P', S\rangle$$

   where

   $$P' = \{\ S \rightarrow f_1[C, A, B], \quad A \rightarrow f_2[A], \quad B \rightarrow f_3[B], \quad A \rightarrow f_4[\,], \quad B \rightarrow f_5[\,], \quad C \rightarrow f_6[\,]\ \}$$

   and

   $$\begin{aligned} F = \{ \quad & f_1[\langle Z_1, Z_2\rangle, \langle X_1, X_2, X_3\rangle, \langle Y_1, Y_2\rangle] = \langle Z_1X_1Y_1X_2Y_2X_3Z_2\rangle \\ & f_2[\langle X_1, X_2, X_3\rangle] = \langle aX_1, bX_2, eX_3\rangle \\ & f_2[\langle X_1, X_2\rangle] = \langle dX_1, dX_2\rangle \\ & f_4[\,] = \langle a, b, e\rangle \\ & f_5[\,] = \langle d, d\rangle \\ & f_6[\,] = \langle cc, cc\rangle \quad \} \end{aligned}$$

Question 2

Give an LCFRS in the sRCG format (the format from the previous exercise) for the language

$$L = \{www^Rw^R \mid w \in \{a, b\}^*, w^R \text{ is } w \text{ in reverse order}\}$$

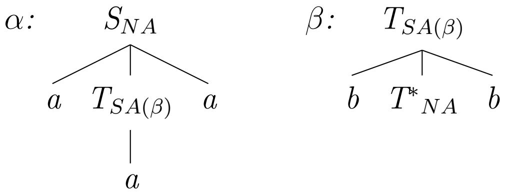
Solution:

$$G = \langle \{A, S\}, \{a, b\}, \{W, X, Y, Z\}, P, S \rangle$$

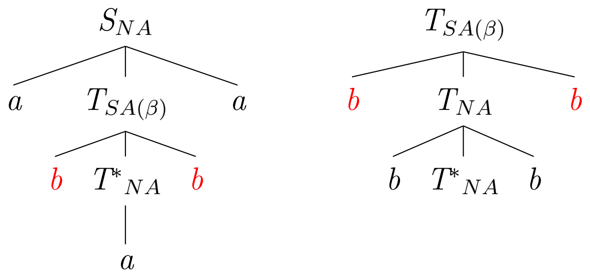
where
$P = \{ \quad S(WXYZ) \to A(W, X, Y, Z), \quad A(\varepsilon, \varepsilon, \varepsilon, \varepsilon) \to \varepsilon,$
$\quad A(aW, aX, Ya, Za) \to A(W, X, Y, Z), \quad A(bW, bX, Yb, Zb) \to A(W, X, Y, Z) \quad \}$

Alternative solution with fan-out 3:
$P = \{ \quad S(WXZ) \to A(W, X, Z), \quad A(\varepsilon, \varepsilon, \varepsilon) \to \varepsilon,$
$\quad A(aW, aXa, Za) \to A(W, X, Z), \quad A(bW, bXb, Zb) \to A(W, X, Z) \quad \}$

Question 3 *Consider the following TAG:*



Adjoining at a node v means adding a tuple of 2 components, one to the left and the other to the right of all daughters of the adjunction site:



With respect to α, its span can be aaa or aX_1aX_2a where in the latter case, the tuple $\langle X_1, X_2 \rangle$ was added by adjoining β, i.e., it is an element in the yield of β. Similarly, in β, we can either not adjoin (leads to tuple $\langle b, b \rangle$) or adjoin at the root.

Give a corresponding LCFRS with nonterminals $\{\alpha, \beta\}$, terminals $\{a, b\}$, start symbol α and rules that do exactly these concatenations.

Rules: $\alpha(aaa) \to \varepsilon$, $\alpha(aX_1aX_2a) \to \beta(X_1, X_2)$, $\beta(b, b) \to \varepsilon$, $\beta(X_1b, bX_2) \to \beta(X_1, X_2)$.