Parsing Beyond CFG

Homework 7: LCFRS — Formal Properties and Normal Forms

Laura Kallmeyer, Tatiana Bladier

Sommersemester 2018

Question 1

Show that the language L is not a 2-MCFL:

$$L = \{a^n b^m a^n b^m a^n b^m a^n b^m a^n b^m\}$$

Hint: show that this language does not satisfy the pumping lemma for 2-MCFL.

Solution:

According to the pumping lemma for 2-MCFL, there must be at least one word in the language of the form $w_1 v_1 w_2 v_2 w_3 v_3 w_4 v_4 w_5 v_5$ where $v_1 v_2 v_3 v_4 v_5 \neq \epsilon$ such that the v_i $(1 \leq i \leq 4)$ can be iterated. Each of the $v_1, \dots, v_4$ must necessarily contain either only as or only bs, otherwise the next iteration step would lead to a word outside the language. However, this means that by these iterations only some and not all of the exponents n and m can get increased (since maximally four substrings are iterated but we have five exponents n and five exponents m). I.e., after the next iteration we necessarily obtain a word with either two a-sequences of different length or two b-sequences of different length. This means that the word we obtain by iteration is not in L. Therefore, L does not satisfy the pumping lemma for 2-MCFL and thus is not a 2-MCFL.

Question 2

Consider the following LCFRS in simple RCG format:

$$G = \langle \{A, B, C, D, S\}, \{a, b, c, d\}, \{W, X, Y, Z\}, P, S \rangle$$

where

$$\begin{aligned} P = \{ \quad & S(ZYXW) \to A(W, Y) B(X, Z), \\ & S(XY) \to C(X, Y) \\ & A(\epsilon, b) \to \epsilon, \\ & A(aX, bY) \to A(X, Y), \\ & B(c, \epsilon) \to \epsilon \\ & C(Xc, Yc) \to C(X, Y) \\ & D(d) \to \epsilon \quad \} \end{aligned}$$

1. *Perform the following transformations on this LCFRS in the simple RCG format while obtaining always weakly equivalent LCFRS:*

   (a) *Transform this grammar G into a weakly equivalent ordered LCFRS.*

   (b) *Remove useless rules.*

   (c) *Remove ϵ-rules.*

2. *What is the string language generated by this grammar?*

Solution:

1. Simplifying the LCFRS G includes the following steps:

(a) The only problematic rule is $S(ZYXW) \rightarrow A(W,Y)B(X,Z)$.
It transforms into $S^(ZYXW) \rightarrow A^{\langle 2,1 \rangle}(Y,W)B^{\langle 2,1 \rangle}(Z,X)$ (if the superscript is the identity, we omit it).

We add $B^{\langle 2,1 \rangle}(\epsilon, c) \rightarrow \epsilon$, $A^{\langle 2,1 \rangle}(b, \epsilon) \rightarrow \epsilon$, and $A^{\langle 2,1 \rangle}(bY, aX) \rightarrow A(X,Y)$

Now, there is again a problematic rule $A^{\langle 2,1 \rangle}(bY, aX) \rightarrow A(X,Y)$. We transform it into $A^{\langle 2,1 \rangle}(bY, aX) \rightarrow A^{\langle 2,1 \rangle}(Y,X)$

In the following, for reasons of readability, we replace $A^{\langle 2,1 \rangle}$ and $B^{\langle 2,1 \rangle}$ with new symbols E and F, respectively.

Resulting grammar after the first step has the following look:

$$G = \langle \{A, C, D, E, F, S\}, \{a, b, c, d\}, \{W, X, Y, Z\}, P, S \rangle$$

where

$$\begin{aligned} P = \{ \quad & S(ZYXW) \rightarrow E(Y,W)F(Z,X), \\ & S(XY) \rightarrow C(X,Y) \\ & E(b,\epsilon) \rightarrow \epsilon, \\ & E(bY, aX) \rightarrow E(Y,X), \\ & F(\epsilon, c) \rightarrow \epsilon \\ & C(Xc, Yc) \rightarrow C(X,Y) \\ & D(d) \rightarrow \epsilon \qquad \} \end{aligned}$$

(b) We remove the rules $S(XY) \rightarrow C(X,Y)$, $C(Xc, Yc) \rightarrow C(X,Y)$, and $D(d) \rightarrow \epsilon$, since we cannot generate any spans of terminals with these rules.

Resulting grammar after the second step has the following look:

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

where

$$\begin{aligned} P = \{ \quad & S(ZYXW) \rightarrow E(Y,W)F(Z,X), \\ & E(b,\epsilon) \rightarrow \epsilon, \\ & E(bY, aX) \rightarrow E(Y,X), \\ & F(\epsilon, c) \rightarrow \epsilon \end{aligned}$$

(c) Remove ϵ-rules.

Resulting set of non-terminals $N_\epsilon = \{S^1, E^{10}, E^{11}, F^{01}\}$ and the new productions:

$$\begin{aligned} P = \{ \quad & S^1(YXW) \rightarrow E^{11}(Y,W)F^{01}(X), \\ & S^1(XW) \rightarrow E^{10}(W)F^{01}(X), \\ & E^{11}(bY, aX) \rightarrow E^{11}(Y,X), \\ & E^{11}(bY, a) \rightarrow E^{10}(Y), \\ & E^{10}(b) \rightarrow \epsilon, \\ & F^{01}(c) \rightarrow \epsilon \qquad \} \end{aligned}$$

2. $L = \{b^n c a^{n-1} \mid n \geq 1\}$