# Harvard University
## Computer Science 121
## Problem Set 5

Due Tuesday, October 22, 2013 at 11:59 PM.
Submit your solutions electronically to cs121+ps5@seas.harvard.edu with "ps5 submission" in the subject line. The solutions to each part should be attached as separate PDF files, called lastname+ps5a.pdf, lastname+ps5b.pdf, and lastname+ps5c.pdf.
Late problem sets may be turned in until Friday, October 25, 2013 at 11:59 PM with a 20% penalty.
See syllabus for collaboration policy.

## PART C (Graded by Nick)

PROBLEM 1 (15 points)

Let ROMANIAN denote the set of all grammatical sentences in Romanian over an alphabet $\Sigma$ consisting of all words in Romanian. Let $S$ denote the set of all Romanian sentenes as defined below:

$$S = \{(\text{Pe cine})^i\ (\text{cui})^j \text{ ai vrut } (\text{sa rogi})^i\ (\text{sa spuna})^j \text{ povestea: } i, j > 0\}$$

Let $f : \{a, b, c, d, x, y\}^* \to \Sigma^*$ be the following homomorphism:

$$f(w) = \begin{cases} \text{pe cine} & \text{if } w = a \\ \text{cui} & \text{if } w = b \\ \text{sa rogi} & \text{if } w = c \\ \text{sa spuna} & \text{if } w = d \\ \text{ai vrut} & \text{if } w = x \\ \text{povestea} & \text{if } w = y \end{cases}$$

(A) Given that a string "(Pe cine)$^{i_0}$ (cui)$^{j_0}$ ai vrut (sa rogi)$^{i_1}$ (sa spuna)$^{j_1}$ povestea" $\in \Sigma^*$ is in ROMANIAN if and only if $i_0 = i_1 > 0$ and $j_0 = j_1 > 0$, describe the following language $L$ in set notation:

$$L = f^{-1}(\text{ROMANIAN} \cap ((\text{pe cine})^*(\text{cui})^* \text{ ai vrut } (\text{sa rogi})^*(\text{sa spuna})^* \text{ povestea}))$$

(B) Using $L$, argue that ROMANIAN is non-context-free.

(Note: For those interested, we give the translation of the sentences of $S$ below. The top line is the Romanian, the middle line is a word for word translation, and the bottom line is the English translation.)

(1) (Pe cine)$^i$ (cui)$^j$ ai vrut (sa rogi)$^i$ (sa spuna)$^j$ povestea?
(who.ACC)$^i$ (who.DAT)$^j$ have wanted (to ask)$^i$ (to tell)$^j$ story?
"Who have you wanted to ask to ask who ... to tell who... to tell who the story?"

(A)

$$f^{-1}(\text{ROMANIAN} \cap ((\text{pe cine})^*(\text{cui})^* \text{ ai vrut } (\text{sa rogi})^*(\text{sa spuna})^* \text{ povestea})) = \{a^n b^m c^n d^m : n, m > 0\}$$

(B) We prove the language $L$ as defined above is non-context-free. The closure of context free languages under inverse-homomorphism and intersection with regular languages thus ensures the non-context-freeness of Romanian.

Let $p$ be the pumping length associated with $L$. Let $w = a^p b^p c^p d^p$. We consider the variuous cases for partition $w = uvxyz$.

1. $v, y$ contain only $a$'s: Take $w' = uv^2xy^2z$. Then $w'$ has more $a$'s than $c$'s, so $w' \notin L$.
2. $v$ contains only $a$'s and $y$ contains only $b$'s; or $v$ or $y$ contain $a$'s and $b$'s: In this case $w' = uv^2xy^2z$ will have more $a$'s and $b$'s than $c$'s and $d$'s, so $w' \notin L$.
3. $v, y$ contain only $b$'s: See 1. above for $v, y$ contain only $a$'s.
4. $v$ contains only $b$'s and $y$ contains only $c$'s; or $v$ or $y$ contain $b$'s and $c$'s: See 2. above, identical argument.
5. $v, y$ contain only $c$'s: See 1. above for $v, y$ contain only $a$'s.
6. $v$ contains only $c$'s and $y$ contains only $d$'s; or $v$ or $y$ contain $c$'s and $d$'s: See 2. above, identical argument.
7. $v, y$ contain only $d$'s: See 1. above for $v, y$ contain only $a$'s.

It follows that $L$ is non-context-free, thus completing the proof.

## PROBLEM 2 (5+5 points)

Let $G = (V, \Sigma, R, S)$ where $V = \{S, V\}$, $\Sigma = \{a, b\}$, and $R$ is the set of rules:

$$\begin{aligned} S &\rightarrow bSS \mid aS \mid aV \\ V &\rightarrow aVb \mid bVa \mid VV \mid \varepsilon \end{aligned}$$

(A) Transform $G$ into an equivalent grammar $G'$ in Chomsky normal form.

(B) Verify that the string $abaab$ is generated by $G'$, using the recognition algorithm for grammars in Chomsky normal form given in class. Show the complete filled-in matrix.

(A) Add a new start symbol:
$S' \rightarrow S$
$S \rightarrow bSS \mid aS \mid aV$

$V \to aVb \mid bVa \mid VV \mid \varepsilon$

Eliminate $V \to \varepsilon$:

$S' \to S$
$S \to bSS \mid aS \mid aV \mid a$
$V \to aVb \mid bVa \mid VV \mid ab \mid ba$

Eliminate long rules:

$S' \to S$
$S \to bW \mid aS \mid aV \mid a$
$V \to aX \mid bY \mid VV \mid ab \mid ba$
$W \to SS$
$X \to Vb$
$Y \to Va$

Eliminate unit rules:

$S' \to bW \mid aS \mid aV \mid a$
$S \to bW \mid aS \mid aV \mid a$
$V \to aX \mid bY \mid VV \mid ab \mid ba$
$W \to SS$
$X \to Vb$
$Y \to Va$

Eliminate terminal-generating rules:

$S' \to BW \mid AS \mid AV \mid a$
$S \to BW \mid AS \mid AV \mid a$
$V \to AX \mid BY \mid VV \mid AB \mid BA$
$W \to SS$
$X \to VB$
$Y \to VA$
$A \to a$
$B \to b$

(B)

$S'$ is present in the lower left corner, so $abaab$ is generated by $G'$.

```
       a
|S'SA|      b
|V   |B   |      a
|S'SY|V   |S'SA|      a
|S'SW|S'SY|S'SW|S'SA|      b
|S'S |V   |S'S |V   |B   |
```