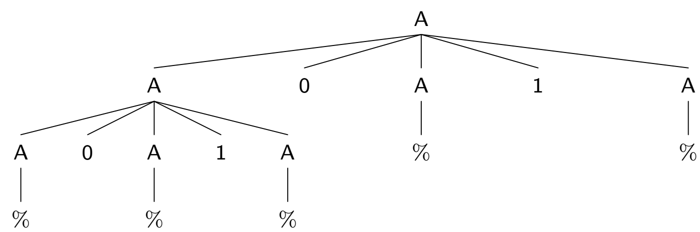
CS 591 S2—Formal Language Theory: Integrating Experimentation and Proof—Fall 2018

Problem Set 6

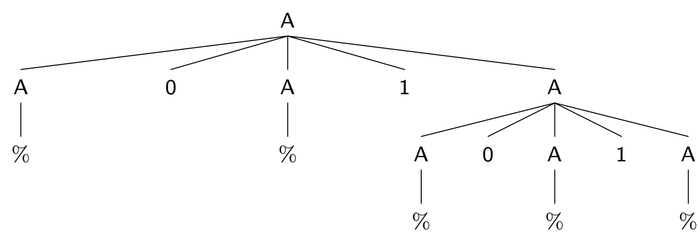
Model Answers

Problem 1

Let pt_1 be the parse tree



And let pt_2 be the parse tree



To check that our answer is correct, we proceed as follows:

```
- val gram = Gram.input "";
@ {variables} A {start variable} A
@ {productions} A -> % | A0A1A
@ .
val gram = - : gram
- val pt1 = PT.input "";
@ A(A(A(%), 0, A(%), 1, A(%)), 0, A(%), 1, A(%))
@ .
val pt1 = - : pt
- val pt2 = PT.input "";
@ A(A(%), 0, A(%), 1, A(A(%), 0, A(%), 1, A(%)))
@ .
val pt2 = - : pt
- PT.equal(pt1, pt2);
val it = false : bool
- Gram.validPT gram pt1;
val it = true : bool
```

```
- Sym.equal(Gram.startVariable gram, PT.rootLabel pt1);
val it = true : bool
- val x1 = PT.yield pt1;
val x1 = [-,-,-,-] : str
- Gram.validPT gram pt2;
val it = true : bool
- Sym.equal(Gram.startVariable gram, PT.rootLabel pt2);
val it = true : bool
- val x2 = PT.yield pt2;
val x2 = [-,-,-,-] : str
- Str.equal(x1, x2);
val it = true : bool
- SymSet.subset(Str.alphabet x1, Gram.alphabet gram);
val it = true : bool
```

Problem 2

(a) Suppose, toward a contradiction, that X is regular. Thus there is an $n \in \mathbb{N}$ with the property of the Pumping Lemma, where X has been substituted for L. Suppose $z = 1^n 3^n$. Because $n \leq n+0 \leq 2n$, we have that $z = 1^n 3^n = 1^n 2^0 3^n \in X$. Thus, since $|z| = 2n \geq n$, it follows that there are $u, v, w \in \textbf{Str}$ such that $z = uvw$ and properties (1)–(3) of the lemma hold. Since $uvw = z = 1^n 3^n$, (1) tells us that there are $i, j, k \in \mathbb{N}$ such that

$$u = 1^i, \quad v = 1^j, \quad w = 1^k 3^n, \quad i + j + k = n.$$

By (2), we have that $j \geq 1$, and thus that $i + k = n - j < n$. By (3), we have that $1^{i+k} 2^0 3^n = 1^i 1^k 3^n = uw = uv^0 w \in X$. Thus $n \leq (i+k)+0 \leq 2n$, so that $n \leq i+k$. But $i+k < n$—contradiction. Thus X is not regular.

(b) G is the grammar

$$\mathsf{A} \rightarrow \mathsf{B} \mid \mathsf{12B3} \mid \mathsf{1A3} \mid \mathsf{11A3},$$
$$\mathsf{B} \rightarrow \% \mid \mathsf{2B3} \mid \mathsf{22B3}.$$

(c) First we put the Forlan syntax

```
{variables} A, B {start variable} A
{productions}
A -> B | 12B3 | 1A3 | 11A3;
B -> % | 2B3 | 22B3
```

for G in the file `ps6-p2-gram.txt`. Next, we put the following testing code in the file `ps6-p2-testing.sml`:

```
(* the symbols 1, 2 and 3 *)

val one   : sym = Sym.fromString "1"
```

```
val two   : sym = Sym.fromString "2"
val three : sym = Sym.fromString "3"

(* the alphabet {1, 2, 3} *)

val syms123 = SymSet.fromString "1, 2, 3"

(* the language {1, 2, 3} *)

val strs123 = StrSet.map (fn a => [a]) syms123

(* numConseq(a, x) returns (n, y), where n is the length of the
   longest prefix of the str x all of whose elements are the symbol a,
   and the str y is the result of removing that prefix from x *)

fun numConseq(a, x) =
      let fun num(n, nil)     = (n, nil)
            | num(n, b :: bs) =
                if Sym.equal(b, a)
                then num(n + 1, bs)
                else (n, b :: bs)
      in num(0, x) end

(* in_X x tests whether the str x is in the language X *)

fun in_X x =
      let val (i, u) = numConseq(one, x)
          val (j, v) = numConseq(two, u)
          val (k, w) = numConseq(three, v)
      in null w andalso k <= i + j andalso i + j <= 2 * k end

(* if n >= 0, then upto n returns the set of all strs over the
   alphabet {1, 2, 3} of length <= n *)

fun upto 0 = StrSet.fromString "%"
  | upto n = StrSet.union(StrSet.power(strs123, n), upto(n - 1))

(* if n >= 0, boundedTest n gram checks that the alphabet of gram is
   {1, 2, 3}, and assesses gram using all test data of length <= n *)

fun boundedTest n gram =
      let val alls   = upto n
          val goods  = Set.filter in_X alls
          val bads   = StrSet.minus(alls, goods)
          val gen    = Gram.generated gram
          val notGen = not o gen
      in SymSet.equal(Gram.alphabet gram, syms123) andalso
         Set.all gen goods andalso Set.all notGen bads
```

```
    end

(* test gram checks that the alphabet of gram is {1, 2, 3}, and
   assesses gram using all test data of length <= 10 *)

val test = boundedTest 10
```

Then we invoke Forlan and proceed as follows:

```
- val gram = Gram.input "ps6-p2-gram.txt";
val gram = - : gram
- use "ps6-p2-testing.sml";
[opening ps6-p2-testing.sml]
val one = - : sym
val two = - : sym
val three = - : sym
val syms123 = - : sym set
val strs123 = - : str set
val numConseq = fn : sym * sym list -> int * sym list
val in_X = fn : sym list -> bool
val upto = fn : int -> str set
val boundedTest = fn : int -> gram -> bool
val test = fn : gram -> bool
val it = () : unit
- test gram;
val it = true : bool
```

(d) Clearly, **alphabet** $G = \{1, 2, 3\}$. Let $Y = \{\, 2^j 3^k \mid j, k \in \mathbb{N} \text{ and } k \leq j \leq 2k \,\}$.

Lemma PS6.2.1

(A) For all $w \in \Pi_{\mathsf{A}}$, $w \in X$.

(B) For all $w \in \Pi_{\mathsf{B}}$, $w \in Y$.

Proof. By induction on Π. There are seven productions to consider.

($\mathsf{A} \to \mathsf{B}$) Suppose $w \in \Pi_{\mathsf{B}}$, and assume the inductive hypothesis: $w \in Y$. We must show that $w \in X$. Because $w \in Y$, we have that $w = 2^j 3^k$ for some $j, k \in \mathbb{N}$ such that $k \leq j \leq 2k$. Thus $w = 1^0 2^j 3^k$ and $k \leq 0 + j \leq 2k$, showing that $w \in X$.

($\mathsf{A} \to \mathsf{12B3}$) Suppose $w \in \Pi_{\mathsf{B}}$, and assume the inductive hypothesis: $w \in Y$. We must show that $12w3 \in X$. Because $w \in Y$, we have that $w = 2^j 3^k$ for some $j, k \in \mathbb{N}$ such that $k \leq j \leq 2k$. Thus $12w3 = 122^j 3^k 3 = 1^1 2^{j+1} 3^{k+1}$. Because $k \leq j$, we have that $k + 1 \leq j + 1 \leq 1 + (j + 1)$. Because $j \leq 2k$, we have that $1 + (j+1) = j + 2 \leq 2k + 2 = 2(k+1)$. Thus $k + 1 \leq 1 + (j + 1) \leq 2(k + 1)$, completing the proof that $12w3 \in X$.

($\mathsf{A} \to \mathsf{1A3}$) Suppose $w \in \Pi_{\mathsf{A}}$, and assume the inductive hypothesis: $w \in X$. We must show that $1w3 \in X$. Because $w \in X$, we have that $w = 1^i 2^j 3^k$ for some $i, j, k \in \mathbb{N}$ such that $k \leq i + j \leq 2k$. Thus $1w3 = 11^i 2^j 3^k 3 = 1^{i+1} 2^j 3^{k+1}$. Because $k \leq i + j$, we have that

$k+1 \leq (i+1)+j$. Because $i+j \leq 2k$, we have that $(i+1)+j \leq 2k+1 \leq 2k+2 = 2(k+1)$. Thus $k+1 \leq (i+1)+j \leq 2(k+1)$, completing the proof that $1w3 \in X$.

($\mathsf{A} \to \mathsf{11A3}$) Suppose $w \in \Pi_{\mathsf{A}}$, and assume the inductive hypothesis: $w \in X$. We must show that $11w3 \in X$. Because $w \in X$, we have that $w = 1^i 2^j 3^k$ for some $i, j, k \in \mathbb{N}$ such that $k \leq i+j \leq 2k$. Thus $11w3 = 111^i 2^j 3^k 3 = 1^{i+2} 2^j 3^{k+1}$. Because $k \leq i+j$, we have that $k+1 \leq (i+j)+1 \leq (i+2)+j$. Because $i+j \leq 2k$, we have that $(i+2)+j \leq 2k+2 = 2(k+1)$. Thus $k+1 \leq (i+2)+j \leq 2(k+1)$, completing the proof that $11w3 \in X$.

($\mathsf{B} \to \%$) We must show that $\% \in Y$. And this follows since $\% = 2^0 3^0$ and $0 \leq 0 \leq 2 * 0$.

($\mathsf{B} \to \mathsf{2B3}$) Suppose $w \in \Pi_{\mathsf{B}}$, and assume the inductive hypothesis: $w \in Y$. We must show that $2w3 \in Y$. Because $w \in Y$, we have that $w = 2^j 3^k$ for some $j, k \in \mathbb{N}$ such that $k \leq j \leq 2k$. Thus $2w3 = 22^j 3^k 3 = 2^{j+1} 3^{k+1}$ and $k+1 \leq j+1 \leq 2k+1 \leq 2k+2 = 2(k+1)$, completing the proof that $2w3 \in Y$.

($\mathsf{B} \to \mathsf{22B3}$) Suppose $w \in \Pi_{\mathsf{B}}$, and assume the inductive hypothesis: $w \in Y$. We must show that $22w3 \in Y$. Because $w \in Y$, we have that $w = 2^j 3^k$ for some $j, k \in \mathbb{N}$ such that $k \leq j \leq 2k$. Thus $22w3 = 222^j 3^k 3 = 2^{j+2} 3^{k+1}$ and $k+1 \leq j+2 \leq 2k+2 = 2(k+1)$, completing the proof that $22w3 \in Y$.

□

Lemma PS6.2.2
$Y \subseteq \Pi_{\mathsf{B}}$.

Proof. It will suffice to show that, for all $k \in \mathbb{N}$, for all $j \in \mathbb{N}$, if $k \leq j \leq 2k$, then $2^j 3^k \in \Pi_{\mathsf{B}}$. We proceed by mathematical induction.

(basis step) We must show that, for all $j \in \mathbb{N}$, if $0 \leq j \leq 2 * 0$, then $2^j 3^0 \in \Pi_{\mathsf{B}}$. Suppose $j \in \mathbb{N}$ and $0 \leq j \leq 2 * 0$. Thus $j = 0$, so that $2^j 3^0 = 2^0 3^0 = \% \in \Pi_{\mathsf{B}}$, because of the production $\mathsf{B} \to \%$.

(inductive step) Suppose $k \in \mathbb{N}$, and assume the inductive hypothesis: for all $j \in \mathbb{N}$, if $k \leq j \leq 2k$, then $2^j 3^k \in \Pi_{\mathsf{B}}$. We must show that, for all $j \in \mathbb{N}$, if $k+1 \leq j \leq 2(k+1)$, then $2^j 3^{k+1} \in \Pi_{\mathsf{B}}$. Suppose $j \in \mathbb{N}$ and $k+1 \leq j \leq 2(k+1)$. We must show that $2^j 3^{k+1} \in \Pi_{\mathsf{B}}$. Because $j \leq 2(k+1) = 2k+2$, we have that $j-1 \leq 2k+1$. There are two cases to consider.

- Suppose $j-1 \leq 2k$. Since $k+1 \leq j$, we have that $k \leq j-1 \leq 2k$. Hence the inductive hypothesis tells us that $2^{j-1} 3^k \in \Pi_{\mathsf{B}}$. Thus $2^j 3^{k+1} = 22^{j-1} 3^k 3 \in \Pi_{\mathsf{B}}$, because of the production $\mathsf{B} \to \mathsf{2B3}$.
- Suppose $j-1 = 2k+1$. Then $j-2 = 2k$. And $k \leq 2k = j-2$, so that $k \leq j-2 \leq 2k$. Hence the inductive hypothesis tells us that $2^{j-2} 3^k \in \Pi_{\mathsf{B}}$. Thus $2^j 3^{k+1} = 222^{j-2} 3^k 3 \in \Pi_{\mathsf{B}}$, because of the production $\mathsf{B} \to \mathsf{22B3}$.

□

Lemma PS6.2.3
$X \subseteq \Pi_{\mathsf{A}}$.

Proof. It will suffice to show that, for all $k \in \mathbb{N}$, for all $i, j \in \mathbb{N}$, if $k \leq i+j \leq 2k$, then $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{A}$. We proceed by mathematical induction.

(basis step) We must show that, for all $i, j \in \mathbb{N}$, if $0 \leq i + j \leq 2 * 0$, then $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^0 \in \Pi_\mathsf{A}$. Suppose $i, j \in \mathbb{N}$ and $0 \leq i + j \leq 2 * 0$. Because $i, j \in \mathbb{N}$, it follows that $i = 0$ and $j = 0$, so that $0 \leq j \leq 2 * 0$, and thus $\mathsf{2}^j\mathsf{3}^0 \in Y$. By Lemma PS6.2.2, it follows that $\mathsf{2}^j\mathsf{3}^0 \in \Pi_\mathsf{B}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^0 = \mathsf{1}^0\mathsf{2}^j\mathsf{3}^0 = \mathsf{2}^j\mathsf{3}^0 \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{B}$.

(inductive step) Suppose $k \in \mathbb{N}$, and assume the inductive hypothesis: for all $i, j \in \mathbb{N}$, if $k \leq i+j \leq 2k$, then $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{A}$. We must show that, for all $i, j \in \mathbb{N}$, if $k+1 \leq i+j \leq 2(k+1)$, then $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} \in \Pi_\mathsf{A}$. Suppose $i, j \in \mathbb{N}$ and $k + 1 \leq i + j \leq 2(k + 1)$. We must show that $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} \in \Pi_\mathsf{A}$. There are three cases to consider.

- Suppose $i = 0$. Since $k + 1 \leq i + j \leq 2(k + 1)$, we have that $k + 1 \leq j \leq 2(k + 1)$, and thus $\mathsf{2}^j\mathsf{3}^{k+1} \in Y$. Hence Lemma PS6.2.2 tells us that $\mathsf{2}^j\mathsf{3}^{k+1} \in \Pi_\mathsf{B}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{1}^0\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{2}^j\mathsf{3}^{k+1} \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{B}$.
- Suppose $i = 1$. Because $k + 1 \leq 1 + j \leq 2(k + 1) = 2k + 2$, we have that $k \leq j \leq 2k + 1$. There are two subcases to consider.
  - Suppose $j \leq 2k$. Then $k \leq j \leq 2k$, so that $\mathsf{2}^j\mathsf{3}^k \in Y$. Hence Lemma PS6.2.2 tells us that $\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{B}$. Because of production $\mathsf{A} \to \mathsf{B}$, we have $\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{A}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{1}\mathsf{2}^j\mathsf{3}^k\mathsf{3} \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{1A3}$.
  - Suppose $j = 2k + 1$. Thus $j - 1 = 2k$. Since $k \leq 2k = j - 1$, it follows that $k \leq j - 1 \leq 2k$, and thus $\mathsf{2}^{j-1}\mathsf{3}^k \in Y$. Hence Lemma PS6.2.2 tells us that $\mathsf{2}^{j-1}\mathsf{3}^k \in \Pi_\mathsf{B}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{1}\mathsf{2}\mathsf{2}^{j-1}\mathsf{3}^k\mathsf{3} \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{12B3}$.
- Suppose $i \geq 2$. Thus $i - 2 \in \mathbb{N}$. Since $k + 1 \leq i + j \leq 2(k + 1) = 2k + 2$, we have that $k + 1 \leq (i - 2) + 2 + j \leq 2k + 2$. Thus $k \leq (i - 2) + 1 + j$ and $(i - 2) + j \leq 2k$. Because $k \leq (i - 2) + 1 + j$, there are two subcases to consider.
  - Suppose $k \leq (i - 2) + j$. Since $k \leq (i - 2) + j \leq 2k$, the inductive hypothesis tells us that $\mathsf{1}^{i-2}\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{A}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{1}\mathsf{1}\mathsf{1}^{i-2}\mathsf{2}^j\mathsf{3}^k\mathsf{3} \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{11A3}$.
  - Suppose $k = (i - 2) + 1 + j$. Thus $k = (i - 1) + j$. Since $(i - 1) + j = k \leq 2k$, we have that $k \leq (i-1)+j \leq 2k$. Hence the inductive hypothesis tells us that $\mathsf{1}^{i-1}\mathsf{2}^j\mathsf{3}^k \in \Pi_\mathsf{A}$. Thus $\mathsf{1}^i\mathsf{2}^j\mathsf{3}^{k+1} = \mathsf{1}\mathsf{1}^{i-1}\mathsf{2}^j\mathsf{3}^k\mathsf{3} \in \Pi_\mathsf{A}$, because of the production $\mathsf{A} \to \mathsf{1A3}$.

□

By Lemmas PS6.2.1(A) and PS6.2.3, we have $\Pi_\mathsf{A} \subseteq X$ and $X \subseteq \Pi_\mathsf{A}$, showing that $\Pi_\mathsf{A} = X$. Thus $L(G) = \Pi_\mathsf{A} = X$.