Section 4.12: The Pumping Lemma for Context-free Languages

Let L be the language $\{ 0^n 1^n 2^n \mid n \in \mathbb{N} \}$.

Question: is L context-free? I.e., is there a grammar that generates L?

Answer:

The LaTeX source of these slides and the associated book, and the distribution of the Forlan toolset are available on the WWW at `http://www.cis.ksu.edu/~allen/forlan/`.

Section 4.12: The Pumping Lemma for Context-free Languages

Let L be the language $\{ 0^n1^n2^n \mid n \in \mathbb{N} \}$.

Question: is L context-free? I.e., is there a grammar that generates L?

Answer: No.

The LaTeX source of these slides and the associated book, and the distribution of the Forlan toolset are available on the WWW at http://www.cis.ksu.edu/~allen/forlan/.

(4.12) Introduction (Cont.)

In this section, we will study the pumping lemma for context-free languages, which can be used to show that many languages are not context-free. We will use the pumping lemma to prove that L is not context-free, and will sketch the proof of the lemma. Building on this result, we'll be able to show that the context-free languages are not closed under intersection or set-difference.

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L,

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L*, there is a* $n \in \mathbb{N}$ *such that,*

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L*, there is a* $n \in \mathbb{N}$ *such that, for all* $z \in \mathbf{Str}$*, if* $z \in L$ *and* $|z| \geq n$*, then*

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L*, there is a* $n \in \mathbb{N}$ *such that, for all* $z \in \mathbf{Str}$*, if* $z \in L$ *and* $|z| \geq n$*, then there are* $u, v, w, x, y \in \mathbf{Str}$ *such that* $z = uvwxy$ *and*

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L*, there is a* $n \in \mathbb{N}$ *such that, for all* $z \in \mathbf{Str}$*, if* $z \in L$ *and* $|z| \geq n$*, then there are* $u, v, w, x, y \in \mathbf{Str}$ *such that* $z = uvwxy$ *and*

(1) $|vwx| \leq n$*;*

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages) *For all context-free languages* L*, there is a* $n \in \mathbb{N}$ *such that, for all* $z \in \mathbf{Str}$*, if* $z \in L$ *and* $|z| \geq n$*, then there are* $u, v, w, x, y \in \mathbf{Str}$ *such that* $z = uvwxy$ *and*

(1) $|vwx| \leq n$*;*

(2) $vx \neq \%$*; and*

(4.12) Statement of Pumping Lemma

Lemma 4.12.1 (Pumping Lemma for Context Free Languages)
For all context-free languages L, there is a $n \in \mathbb{N}$ such that, for all $z \in \mathbf{Str}$, if $z \in L$ and $|z| \geq n$, then there are $u, v, w, x, y \in \mathbf{Str}$ such that $z = uvwxy$ and

(1) $|vwx| \leq n$;

(2) $vx \neq \%$; *and*

(3) $uv^i wx^i y \in L$, *for all* $i \in \mathbb{N}$.

(4.12) Example Use of Pumping Lemma

Before sketching the proof of the pumping lemma, let's see how it can be used to show that $L = \{ 0^n 1^n 2^n \mid n \in \mathbb{N} \}$ is not context-free. Suppose, toward a contradiction that L is context-free. Thus there is an $n \in \mathbb{N}$ with the property of the lemma. Let $z =$

(4.12) Example Use of Pumping Lemma

Before sketching the proof of the pumping lemma, let's see how it can be used to show that $L = \{\, 0^n1^n2^n \mid n \in \mathbb{N} \,\}$ is not context-free. Suppose, toward a contradiction that L is context-free. Thus there is an $n \in \mathbb{N}$ with the property of the lemma. Let $z = 0^n1^n2^n$. Since $z \in L$ and $|z| = 3n \geq n$, we have that there are $u, v, w, x, y \in \mathbf{Str}$ such that $z = uvwxy$ and

(1) $|vwx| \leq n$;

(2) $vx \neq \%$; and

(3) $uv^iwx^iy \in L$, for all $i \in \mathbb{N}$.

Since $0^n1^n2^n = z = uvwxy$, (1) tells us that vwx

(4.12) Example Use of Pumping Lemma

Before sketching the proof of the pumping lemma, let's see how it can be used to show that $L = \{\, 0^n 1^n 2^n \mid n \in \mathbb{N} \,\}$ is not context-free. Suppose, toward a contradiction that L is context-free. Thus there is an $n \in \mathbb{N}$ with the property of the lemma. Let $z = 0^n 1^n 2^n$. Since $z \in L$ and $|z| = 3n \geq n$, we have that there are $u, v, w, x, y \in \mathbf{Str}$ such that $z = uvwxy$ and

(1) $|vwx| \leq n$;

(2) $vx \neq \%$; and

(3) $uv^i wx^i y \in L$, for all $i \in \mathbb{N}$.

Since $0^n 1^n 2^n = z = uvwxy$, (1) tells us that vwx doesn't contain both a 0 and a 2. Thus, either vwx has no 0's, or vwx has no 2's, so that there are two cases to consider.

(4.12) Example (Cont.)

Suppose vwx has no 0's. Thus vx has no 0's. By (2), we have that vx contains a 1 or a 2. Thus uwy:

(4.12) Example (Cont.)

Suppose vwx has no 0's. Thus vx has no 0's. By (2), we have that vx contains a 1 or a 2. Thus uwy:

- has n 0's;

(4.12) Example (Cont.)

Suppose vwx has no 0's. Thus vx has no 0's. By (2), we have that vx contains a 1 or a 2. Thus uwy:

- has n 0's;
- either has less than n 1's or has less than n 2's.

But (3) tells us that

(4.12) Example (Cont.)

Suppose vwx has no 0's. Thus vx has no 0's. By (2), we have that vx contains a 1 or a 2. Thus uwy:

- has n 0's;
- either has less than n 1's or has less than n 2's.

But (3) tells us that $uwy = uv^0wx^0y \in L$, so that uwy has an equal number of 0's, 1's and 2's—contradiction.

The case where vwx has no 2's is similar.

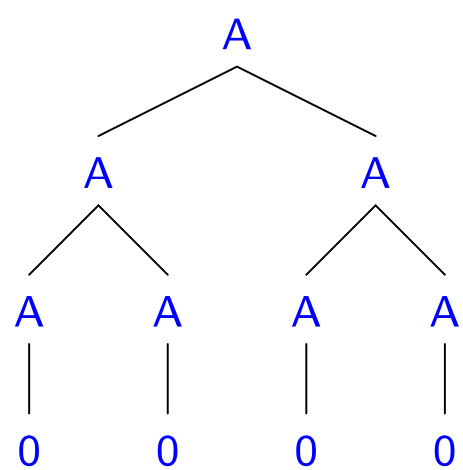
Since we obtained a contradiction in both cases, we have an overall contradiction. Thus L is not context-free.

(4.12) A Fact About CNF Grammars

When we prove the pumping lemma for context-free languages, we will make use of a fact about grammars in Chomsky Normal Form.

Suppose G is a grammar in CNF and that $w \in \mathbf{alphabet}(G)^*$ is the yield of a valid parse tree pt for G whose root label is a variable.

For instance, if G is the grammar with variable A and productions A $\rightarrow$ AA and A $\rightarrow$ 0, then w could be 0000 and pt could be the following tree of height 3:



(4.12) CNF Fact (Cont.)

Generalizing from this example, we can see that if pt has height 3, $|w|$ will never be greater than $4 = 2^2 = 2^{3-1}$.

Question: how can we express an upper bound for $|w|$ as a function of the height k of pt?

Answer: $|w| \leq$

(4.12) CNF Fact (Cont.)

Generalizing from this example, we can see that if pt has height 3, $|w|$ will never be greater than $4 = 2^2 = 2^{3-1}$.

Question: how can we express an upper bound for $|w|$ as a function of the height k of pt?

Answer: $|w| \leq 2^{k-1}$.

We can prove this by induction on pt.

(4.12) Sketch of Proof of Pumping Lemma

Now, let's sketch the proof of the pumping lemma. Suppose L is a context-free language.

(4.12) Sketch of Proof of Pumping Lemma

Now, let's sketch the proof of the pumping lemma. Suppose L is a context-free language. By the results of the preceding section, there is a grammar G in Chomsky Normal Form such that $L(G) = L - \{\%\}$. Let

(4.12) Sketch of Proof of Pumping Lemma

Now, let's sketch the proof of the pumping lemma. Suppose L is a context-free language. By the results of the preceding section, there is a grammar G in Chomsky Normal Form such that $L(G) = L - \{\%\}$. Let $k = |Q_G|$ and $n = 2^k$.

(4.12) Sketch of Proof of Pumping Lemma

Now, let's sketch the proof of the pumping lemma. Suppose L is a context-free language. By the results of the preceding section, there is a grammar G in Chomsky Normal Form such that $L(G) = L - \{\%\}$. Let $k = |Q_G|$ and $n = 2^k$. Suppose $z \in Str$, $z \in L$ and $|z| \geq n$. Since $n \geq 1$, we have that $z \neq \%$. Thus $z \in L - \{\%\} = L(G)$, so that there is a parse tree pt such that pt is valid for G, $\mathbf{rootLabel}(pt) = s_G$ and $\mathbf{yield}(pt) = z$. By our fact about CNF grammars, we have that the height of pt is at least

(4.12) Sketch of Proof of Pumping Lemma

Now, let's sketch the proof of the pumping lemma. Suppose L is a context-free language. By the results of the preceding section, there is a grammar G in Chomsky Normal Form such that $L(G) = L - \{\%\}$. Let $k = |Q_G|$ and $n = 2^k$. Suppose $z \in Str$, $z \in L$ and $|z| \geq n$. Since $n \geq 1$, we have that $z \neq \%$. Thus $z \in L - \{\%\} = L(G)$, so that there is a parse tree pt such that pt is valid for G, $\mathbf{rootLabel}(pt) = s_G$ and $\mathbf{yield}(pt) = z$. By our fact about CNF grammars, we have that the height of pt is at least $k + 1$. (If pt's height were only k, then $|z| \leq 2^{k-1} < n$, which is impossible.) Thus there is a valid path for pt in which we have a repetition of variables; it is this fact that will let us split z into parts u, v, w, x, y with the required properties.

(4.12) Proof Sketch (Cont.)

Instead of doing the rest of the proof in the general case, let's focus on a particular example, so as to understand the idea of the proof.

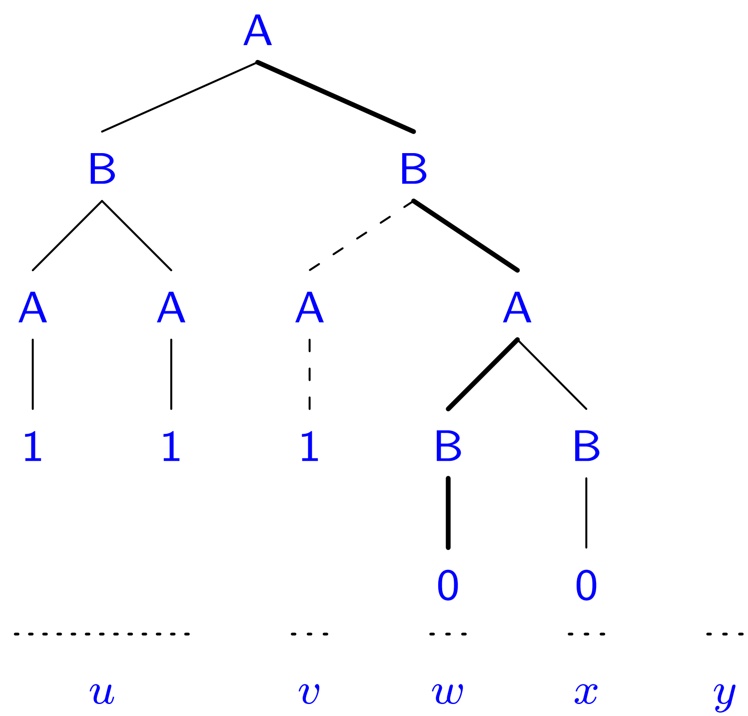
Suppose our grammar G in CNF has variables A and B, start variable A and productions

$$\mathsf{A} \longrightarrow 1 \mid \mathsf{BB},$$
$$\mathsf{B} \longrightarrow 0 \mid \mathsf{AA}.$$

Thus $k = 2$ and $n = 2^k = 2^2 = 4$.

(4.12) Proof Sketch (Cont.)

Suppose $z = 11100$ and let pt be the parse tree:



Then $z \in L(G)$, $|z| = 5 \geq n$ and the height of pt is 4, which is, indeed, at least $k + 1 = 2 + 1 = 3$.

(4.12) Proof Sketch (Cont.)

We continue by selecting a path through pt whose length is equal to the height of the tree: the highlighted path. Since, in the general case, this path has length at least $k + 1$, it contains at least $k + 1$ variables. Thus there is at least one repetition of variables on the path. In the case of our example, both A and B appear twice on our path. Working from the bottom of our path upwards, we look for the first repetition of variables. In the case of our example, this takes us to the subtree with root label B at position $2 \to \textbf{nil}$ in pt. From this node we can also choose a path (indicated by the dashed line) that takes us to a leaf and only intersects the highlighted path at its beginning.

(4.12) Proof Sketch (Cont.)

Let u, v, w, x and y be the indicated parts of the yield of pt ($y = \%$). Thus:

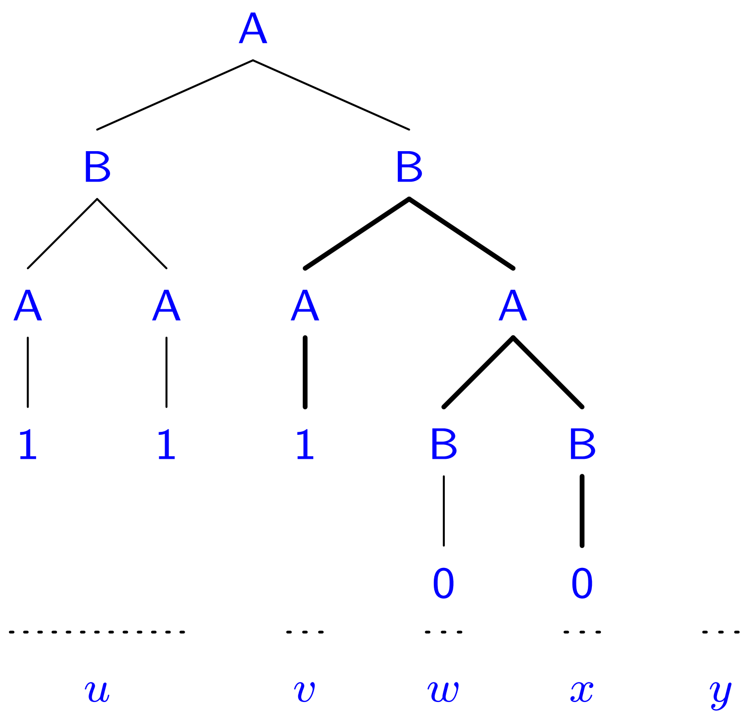
- w is the yield of the subtree with root label B at position $2 \rightarrow 2 \rightarrow 1 \rightarrow \mathbf{nil}$ of pt;
- vwx is the yield of the subtree with root label B at position $2 \rightarrow \mathbf{nil}$ of pt;
- $uvwxy$ is the yield of pt, and so is equal to z.

Because of how we found it, the height of the tree at position $2 \rightarrow \mathbf{nil}$ is no more than $k + 1$, and thus the length of its yield, vwx, is no more than $2^{(k+1)-1} = 2^k = n$. Thus property (1) holds.

In our case, both v and x are non-empty, so that $vx \neq \%$, i.e., property (2) holds. In general, the dashed path leads to a symbol that will be part of either v or x.

(4.12) Proof Sketch (Cont.)

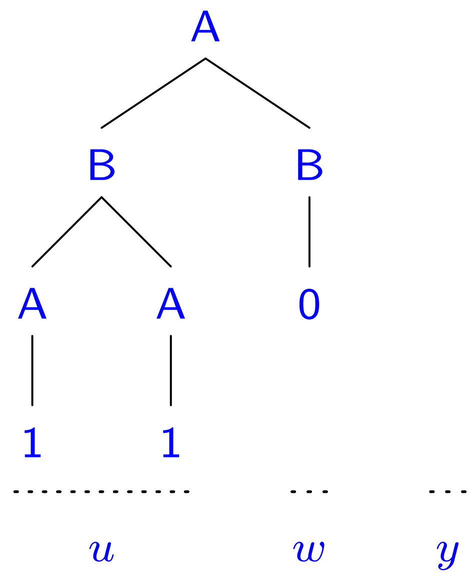
Finally, we must show that property (3) holds, i.e., that $uv^iwx^iy \in L(G) \subseteq L$ for all $i \in \mathbb{N}$. To parse uv^iwx^iy, we can simply repeat the highlighted part of



i times.

(4.12) Proof Sketch (Cont.)

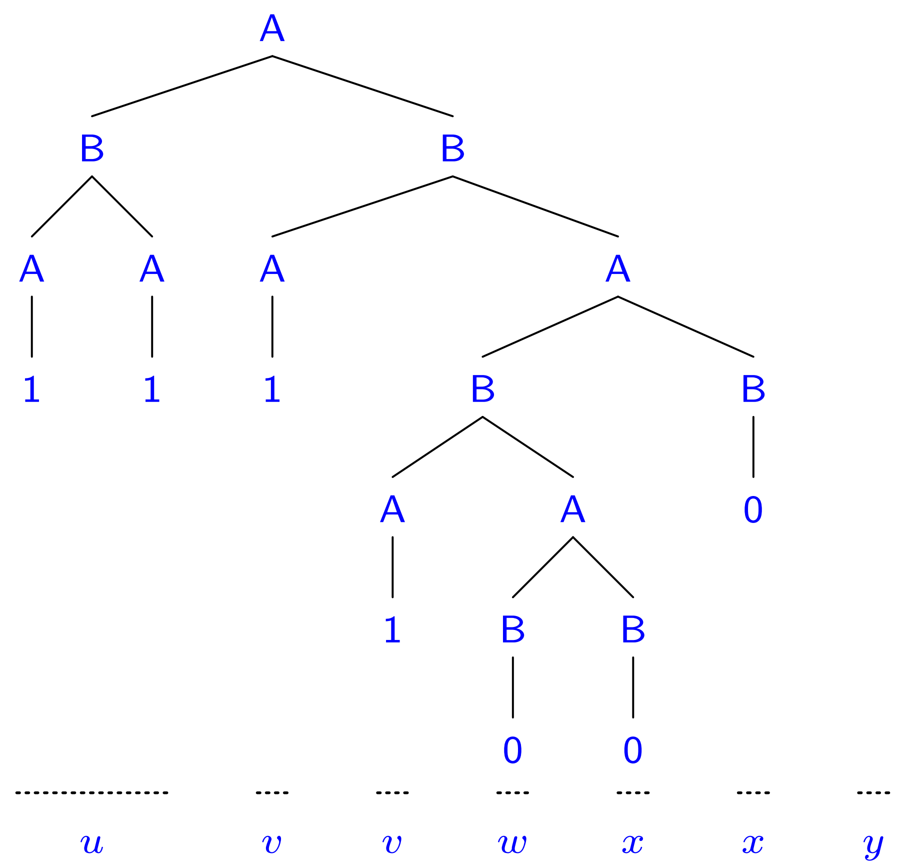
For instance,



is a valid parse tree for G whose yield is $uwy = uv^0wx^0y$.

(4.12) Proof Sketch (Cont.)

And,



is a valid parse tree for G whose yield is $uvvwxxy = uv^2wx^2y$.

(4.12) Consequences of Pumping Lemma

Suppose

$$L = \{\, 0^n 1^n 2^n \mid n \in \mathbb{N} \,\},$$
$$A = \{\, 0^n 1^n 2^m \mid n, m \in \mathbb{N} \,\},$$
$$B = \{\, 0^n 1^m 2^m \mid n, m \in \mathbb{N} \,\}.$$

Of course, L is not context-free.

Question: are A and B context-free?

Answer:

(4.12) Consequences of Pumping Lemma

Suppose

$$L = \{\, 0^n 1^n 2^n \mid n \in \mathbb{N} \,\},$$
$$A = \{\, 0^n 1^n 2^m \mid n, m \in \mathbb{N} \,\},$$
$$B = \{\, 0^n 1^m 2^m \mid n, m \in \mathbb{N} \,\}.$$

Of course, L is not context-free.

Question: are A and B context-free?

Answer: yes, it's easy to find grammars that generate them.

Question: is $A \cap B$ context-free?

Answer:

(4.12) Consequences of Pumping Lemma

Suppose

$$L = \{\, 0^n 1^n 2^n \mid n \in \mathbb{N} \,\},$$
$$A = \{\, 0^n 1^n 2^m \mid n, m \in \mathbb{N} \,\},$$
$$B = \{\, 0^n 1^m 2^m \mid n, m \in \mathbb{N} \,\}.$$

Of course, L is not context-free.

Question: are A and B context-free?

Answer: yes, it's easy to find grammars that generate them.

Question: is $A \cap B$ context-free?

Answer: no—$A \cap B = L$, and L is not context-free.

Thus the context-free languages are not closed under intersection.

(4.12) Consequences (Cont.)

Question: is $\{0, 1, 2\}^* - A$ context-free?

Answer:

(4.12) Consequences (Cont.)

Question: is $\{0, 1, 2\}^* - A$ context-free?

Answer: yes, since this language is the union of the context-free languages

$$\{0, 1, 2\}^* - \{0\}^*\{1\}^*\{2\}^*$$

and

$$\{\, 0^{n_1} 1^{n_2} 2^m \mid n_1, n_2, m \in \mathbb{N} \text{ and } n_1 \neq n_2 \,\},$$

(the first of these languages is regular), and the context-free languages are closed under union.

Similarly, we have that $\{0, 1, 2\}^* - B$ is context-free.

(4.12) Consequences (Cont.)

Let

$$C = (\{0,1,2\}^* - A) \cup (\{0,1,2\}^* - B).$$

Thus C is a context-free subset of $\{0,1,2\}^*$. Since $A, B \subseteq \{0,1,2\}^*$, it is easy to show that

$$\begin{aligned} A \cap B &= \{0,1,2\}^* - ((\{0,1,2\}^* - A) \cup (\{0,1,2\}^* - B)) \\ &= \{0,1,2\}^* - C. \end{aligned}$$

(4.12) Consequences (Cont.)

Let

$$C = (\{0,1,2\}^* - A) \cup (\{0,1,2\}^* - B).$$

Thus C is a context-free subset of $\{0,1,2\}^*$. Since $A, B \subseteq \{0,1,2\}^*$, it is easy to show that

$$\begin{aligned} A \cap B &= \{0,1,2\}^* - ((\{0,1,2\}^* - A) \cup (\{0,1,2\}^* - B)) \\ &= \{0,1,2\}^* - C. \end{aligned}$$

Thus

$$\{0,1,2\}^* - C = A \cap B = L$$

is not context-free. Since $\{0,1,2\}^*$ is regular and thus context-free, it follows that the context-free languages are not closed under set difference.