T065001: Introduction to Formal Languages

Lecture 8: Pushdown automata, context-free languages, grammars (3)

Chapter 2.3 in Sipser's textbook

2025-06-09

(Lecture slides by Yih-Kuen Tsay)

(From Chapter 1.4)

- To understand the power of finite automata we must also understand their limitations.
- Consider the language $B = \{0^n 1^n \mid n \geq 0\}$.
- To recognize B, a machine will have to remember how many 0s have been read so far. This cannot be done with any finite number of states, since the number of 0s is not limited.
- To prove that a language is not regular, we will need a technique based on the *pumping lemma*.

(From Chapter 1.4)

Theorem (1.70)

If A is a regular language, then there is a number p (the pumping length) such that, if s is any string in A and $|s| \geq p$, then s may be divided as $s = xyz$ satisfying:

1. *for each $i \geq 0$, $xy^iz \in A$ (string s can be "pumped"),*
2. *$|y| > 0$, and*
3. *$|xy| \leq p$.*

In Chapter 1.4, we used the pumping lemma above to prove that certain languages are not regular.

Today, we will present an analogous technique for proving that certain languages are not context free.

(From Chapter 1.4)

Theorem (1.70)

If A is a regular language, then there is a number p (the pumping length) such that, if s is any string in A and $|s| \geq p$, then s may be divided as $s = xyz$ satisfying:

1. *for each $i \geq 0$, $xy^iz \in A$ (string s can be "pumped"),*
2. *$|y| > 0$, and*
3. *$|xy| \leq p$.*

In Chapter 1.4, we used the pumping lemma above to prove that certain languages are not regular.

Today, we will present an analogous technique for proving that certain languages are not context free.

(The meaning of "pumped" becomes slightly more complex than before.)

The Pumping Lemma for CFL

Theorem (2.34)

If A is a context-free language, then there is a number p such that, if s is a string in A and $|s| \geq p$, then s may be divided into five pieces, $s = uvxyz$, satisfying the conditions:

1. *for each $i \geq 0$, $uv^ixy^iz \in A$,*
2. *$|vy| > 0$, and*
3. *$|vxy| \leq p$.*

The Pumping Lemma for CFL

Recall: A parse tree visualizes the derivation of a string in a CFG.

Example:

$$\begin{aligned}\langle\text{EXPR}\rangle &\rightarrow \langle\text{EXPR}\rangle + \langle\text{TERM}\rangle \mid \langle\text{TERM}\rangle\\ \langle\text{TERM}\rangle &\rightarrow \langle\text{TERM}\rangle \times \langle\text{FACTOR}\rangle \mid \langle\text{FACTOR}\rangle\\ \langle\text{FACTOR}\rangle &\rightarrow (\langle\text{EXPR}\rangle) \mid \mathtt{a}\end{aligned}$$

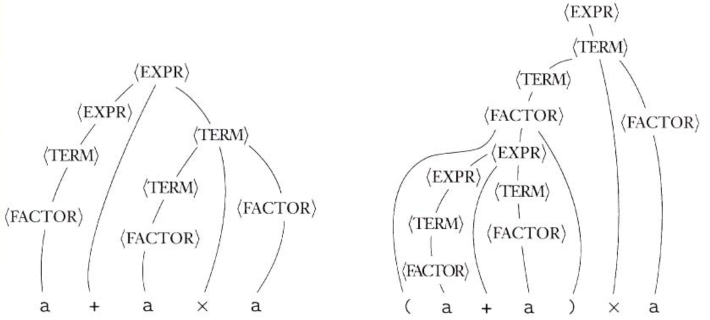

FIGURE 2.5
Parse trees for the strings a+a×a and (a+a)×a

The Pumping Lemma for CFL

Theorem (2.34)

If A is a context-free language, then there is a number p such that, if s is a string in A and $|s| \geq p$, then s may be divided into five pieces, $s = uvxyz$, satisfying the conditions:

1. *for each $i \geq 0$, $uv^ixy^iz \in A$,*
2. *$|vy| > 0$, and*
3. *$|vxy| \leq p$.*

Proof:

- Let G be a CFG that generates A.
- Consider a "sufficiently long" string s in A that satisfies the following condition:
  The parse tree for s is very tall so as to have a long path on which some variable symbol R of G repeats.

The Pumping Lemma for CFL (cont.)

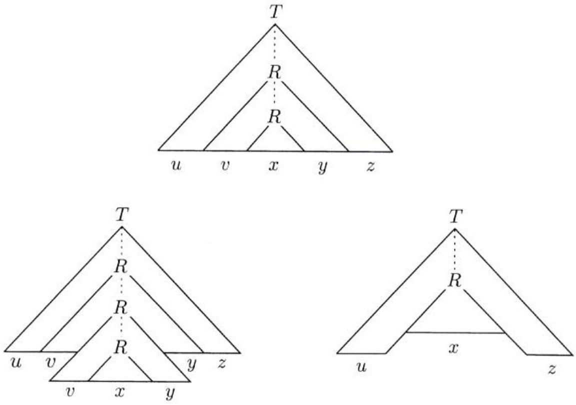

FIGURE **2.35**
Surgery on parse trees

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \longrightarrow w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \rightarrow w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V| + 1$.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V| + 1$. $\Rightarrow \tau$ contains a path π of length $\geq |V|+1$ from the root to a leaf.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \rightarrow w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V|+1$.
  $\Rightarrow$ τ contains a path π of length $\geq |V|+1$ from the root to a leaf.

  But there are only $|V|$ different variables in V. $\Rightarrow$ Some variable R occurs twice on π according to the pigeonhole principle.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V| + 1$. $\Rightarrow \tau$ contains a path π of length $\geq |V|+1$ from the root to a leaf.

  But there are only $|V|$ different variables in V. $\Rightarrow$ Some variable R occurs twice on π according to the pigeonhole principle.

  Divide s into $uvxyz$ as in Fig. 2.35 so that R can generate vxy or x. $\Rightarrow$ Replacing subtrees at R by others gives parse trees for uv^ixy^iz.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V| + 1$. $\Rightarrow$ τ contains a path π of length $\geq |V|+1$ from the root to a leaf.

  But there are only $|V|$ different variables in V. $\Rightarrow$ Some variable R occurs twice on π according to the pigeonhole principle.

  Divide s into $uvxyz$ as in Fig. 2.35 so that R can generate vxy or x. $\Rightarrow$ Replacing subtrees at R by others gives parse trees for uv^ixy^iz.

  - vy cannot be empty, otherwise we'd have a parse tree smaller than τ.

The Pumping Lemma for CFL (cont.)

- Let b be the upper bound on the length of w for any production rule $A \to w$ in G. Every node in a parse tree has $\leq b$ children, so if a parse tree has height h, the length of the generated string is $\leq b^h$.
- Take p to be $b^{|V|+1}$, where V is the set of variables of G.
- Consider any string s with $|s| \geq p$ and let τ be a smallest parse tree for s. If the height of τ is $\leq |V|$ then $|s| \leq b^{|V|}$ by the above, which contradicts $|s| \geq b^{|V|+1}$. Thus, the height of τ is $\geq |V|+1$. $\Rightarrow$ τ contains a path π of length $\geq |V|+1$ from the root to a leaf.

  But there are only $|V|$ different variables in V. $\Rightarrow$ Some variable R occurs twice on π according to the pigeonhole principle.

  Divide s into $uvxyz$ as in Fig. 2.35 so that R can generate vxy or x. $\Rightarrow$ Replacing subtrees at R by others gives parse trees for uv^ixy^iz.
  - vy cannot be empty, otherwise we'd have a parse tree smaller than τ.
  - To ensure $|vxy| \leq p$, choose an R that occurs twice within the bottom $|V|+1$ levels of variables $\Rightarrow$ At most $b^{|V|+1}$ leaves below R.

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

- Let s be $a^p b^p c^p$, where p is the pumping length.

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

- Let s be $a^p b^p c^p$, where p is the pumping length.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

- Let s be $a^p b^p c^p$, where p is the pumping length.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol, e.g.,
     $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x}}^{p}\!\!\overbrace{\underbrace{\cdots}_{y} \cdot b}^{p}\overbrace{c \cdots c}^{p}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in B.

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

- Let s be $a^p b^p c^p$, where p is the pumping length.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol, e.g.,
     $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot\, a}_{}}^{p}\overbrace{b \cdot \underbrace{\cdots}_{y} \cdot b}^{p}\overbrace{c \cdots c}^{p}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in B.
  2. Either v or y contains more than one type of symbol, e.g.,
     $a \cdot \underbrace{\cdots}_{v} \underbrace{\cdots}_{x} \underbrace{\cdot ab \cdot}_{y} \cdots bc \cdots c$, in which case, uv^2xy^2z will have some a's and b's out of order and so is not in B.

Example Non-Context-Free Languages

$B = \{a^n b^n c^n \mid n \geq 0\}$ is not context-free.

Proof: Suppose that B is a CFL. Apply the pumping lemma for CFLs:

- Let s be $a^p b^p c^p$, where p is the pumping length.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol, e.g., $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x}}^{p} \!\!\! \overbrace{\underbrace{\cdots}_{y} \cdot b}^{p} \overbrace{c \cdots c}^{p}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in B.
  2. Either v or y contains more than one type of symbol, e.g., $a \cdot \underbrace{\cdots}_{v} \underbrace{\cdots}_{x} \underbrace{\cdot ab \cdot}_{y} \cdots bc \cdots c$, in which case, uv^2xy^2z will have some a's and b's out of order and so is not in B.

All possible cases result in a contradiction. Thus, B cannot be a CFL.

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Proof: Suppose C is a CFL. It contradicts the pumping lemma because:

- Let s be $a^p b^p c^p$.

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Proof: Suppose C is a CFL. It contradicts the pumping lemma because:

- Let s be $a^p b^p c^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol. Then if no c's in vy:
     $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x}}^{p}$... more precisely $\overbrace{a \cdots a}^{p}\overbrace{b \cdots b}^{p}\overbrace{c \cdots c}^{p}$ with $a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x} \underbrace{\cdots}_{y} \cdot b\, c \cdots c$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in C, or

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Proof: Suppose C is a CFL. It contradicts the pumping lemma because:

- Let s be $a^p b^p c^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol. Then if no c's in vy: $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot\, a}_{}}^{p}$ $\overbrace{b \cdot \cdots \cdot b}^{p}$ $\overbrace{c \cdots c}^{p}$ with $\underbrace{\cdot\, a\, b \cdot}_{x}$, $\underbrace{\cdots}_{y}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in C, or if no a's in vy: $a \cdots ab \cdot \underbrace{\cdots}_{v} \underbrace{\cdot bc \cdot}_{x} \underbrace{\cdots}_{y} \cdot c$, in which case, uv^0xy^0z will have less b's or c's than a's and so is not in C.

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Proof: Suppose C is a CFL. It contradicts the pumping lemma because:

- Let s be $a^p b^p c^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol. Then if no c's in vy:
     $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x}}^{p}\overbrace{\underbrace{\cdots}_{y} \cdot b}^{p}\overbrace{c \cdots c}^{p}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in C, or if no a's in vy:
     $a \cdots ab \cdot \underbrace{\cdots}_{v} \underbrace{\cdot bc \cdot}_{x} \underbrace{\cdots}_{y} \cdot c$, in which case, uv^0xy^0z will have less b's or c's than a's and so is not in C. (If no b's then no a's or no c's.)

Example Non-Context-Free Languages (cont.)

$C = \{a^i b^j c^k \mid 0 \leq i \leq j \leq k\}$ is not context-free.

Proof: Suppose C is a CFL. It contradicts the pumping lemma because:

- Let s be $a^p b^p c^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. Both v and y contain only one type of symbol. Then if no c's in vy:
     $\overbrace{a \cdot \underbrace{\cdots}_{v} \underbrace{\cdot a\, b \cdot}_{x}}^{p}$ ... $\overbrace{b \cdot \underbrace{\cdots}_{y} \cdot b}^{p}\overbrace{c \cdots c}^{p}$, in which case, uv^2xy^2z will have more a's or b's than c's and so is not in C, or if no a's in vy:
     $a \cdots ab \cdot \underbrace{\cdots}_{v} \underbrace{\cdot bc \cdot}_{x} \underbrace{\cdots}_{y} \cdot c$, in which case, uv^0xy^0z will have less b's or c's than a's and so is not in C. (If no b's then no a's or no c's.)
  2. Either v or y contains more than one type of symbol, e.g.,
     $a \cdot \underbrace{\cdots}_{v} \underbrace{\cdots}_{x} \underbrace{\cdot ab \cdot}_{y} \cdots bc \cdots c$, in which case, uv^2xy^2z will have some a's and b's out of order and so is not in C.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

(A failed first attempt to prove that D is not context free.)
Let $s' = 0^p 1 0^p 1$. Then s' belongs to D and has length $\geq p$, as required.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

(A failed first attempt to prove that D is not context free.)
Let $s' = 0^p 1 0^p 1$. Then s' belongs to D and has length $\geq p$, as required.

However, it does not lead to a contradiction when we apply the pumping lemma (see this week's exercises). FAIL!

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

(A failed first attempt to prove that D is not context free.)
Let $s' = 0^p 1 0^p 1$. Then s' belongs to D and has length $\geq p$, as required.

However, it does not lead to a contradiction when we apply the pumping lemma (see this week's exercises). FAIL!

$\Rightarrow$ We need to take some other string instead of s'.

The next page shows a string s that will work.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

- Let s be $0^p 1^p 0^p 1^p$.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

- Let s be $0^p1^p0^p1^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \le p$):
  1. The substring vxy is entirely within the first or second half, e.g., $\overbrace{0\cdots\mathrlap{\underbrace{\cdots0\,1\cdot\cdot}_{vxy}}\cdots0}^{p}\overbrace{\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,
$\overbrace{0\cdots\rlap{$\underbrace{}_{vxy}$}\cdots0}^{p}\overbrace{1\cdots\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,
$\overbrace{0\cdots\cdots 0}^{p}\overbrace{1\cdots\cdots 1}^{p}\overbrace{0\cdots 0}^{p}\overbrace{1\cdots 1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even. Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,

$$\overbrace{0\cdots\cdots 0}^{p}\overbrace{1\cdots\cdots\cdots 1}^{p}\overbrace{0\cdots 0}^{p}\overbrace{1\cdots 1}^{p}$$

with vxy spanning $\underbrace{\cdots 0\,1\cdots}_{vxy}$ across the boundary of the first two blocks, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.

Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.

By conditions 2 and 3 of the pumping lemma, $4p + 2 \leq |t| \leq 5p$.

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g., $\overbrace{0\cdots\cdots 0}^{p}\overbrace{1\cdots\cdots 1}^{p}\overbrace{0\cdots 0}^{p}\overbrace{1\cdots 1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.
Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.

By conditions 2 and 3 of the pumping lemma, $4p + 2 \le |t| \le 5p$.

Now, $|t| \le 5p$ gives $|t| - 3p + 1 \le 2p + 1$

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,

$$\overbrace{0\cdots\cdots 0}^{p}\overbrace{1\cdots\cdots 1}^{p}\overbrace{0\cdots 0}^{p}\overbrace{1\cdots 1}^{p},$$

(with $\underbrace{\cdots 0\,1\cdot}_{vxy}$ spanning the end of the first block and the start of the second) in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.

Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.

By conditions 2 and 3 of the pumping lemma, $4p + 2 \leq |t| \leq 5p$.

Now, $|t| \leq 5p$ gives $|t| - 3p + 1 \leq 2p + 1$ and $|t| \geq 4p + 2$ gives $\frac{|t|}{2} + 1 \geq 2p + 2$;

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g., $\overbrace{0\cdots\cdots0}^{p}\overbrace{1\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.
Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.
By conditions 2 and 3 of the pumping lemma, $4p + 2 \le |t| \le 5p$.
Now, $|t| \le 5p$ gives $|t| - 3p + 1 \le 2p + 1$ and $|t| \ge 4p + 2$ gives $\frac{|t|}{2} + 1 \ge 2p + 2$; combining the inequalities, we get $|t| - 3p + 1 < \frac{|t|}{2} + 1 = b$.

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,
$\overbrace{0\cdots\cdots0}^{p}\overbrace{1\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.
Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.
By conditions 2 and 3 of the pumping lemma, $4p + 2 \leq |t| \leq 5p$.
Now, $|t| \leq 5p$ gives $|t| - 3p + 1 \leq 2p + 1$ and $|t| \geq 4p + 2$ gives $\frac{|t|}{2} + 1 \geq 2p + 2$; combining the inequalities, we get $|t| - 3p + 1 < \frac{|t|}{2} + 1 = b$.
Next, $\frac{|t|}{2} + (2p + 1) \leq \frac{|t|}{2} + \frac{|t|}{2} = |t|$ implies $b = \frac{|t|}{2} + 1 \leq |t| - 2p$.

Example Non-Context-Free Languages (cont.)

Remark: Here is a more formal proof of the statement in case 1.

1. The substring vxy is entirely within the first or second half, e.g.,

$$\overbrace{0\cdots\underbrace{\cdots 0}\;}^{p}\overbrace{\underbrace{1\cdot}_{vxy}\cdots\cdots 1}^{p}\overbrace{0\cdots 0}^{p}\overbrace{1\cdots 1}^{p}$$

, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Write $t = uv^2xy^2z$. If $|t|$ is odd then $t \notin D$. $\Rightarrow$ We may assume that $|t|$ is even.
Let $b = \frac{|t|}{2} + 1$ be the starting position of t's second half.

By conditions 2 and 3 of the pumping lemma, $4p + 2 \leq |t| \leq 5p$.

Now, $|t| \leq 5p$ gives $|t| - 3p + 1 \leq 2p + 1$ and $|t| \geq 4p + 2$ gives $\frac{|t|}{2} + 1 \geq 2p + 2$; combining the inequalities, we get $|t| - 3p + 1 < \frac{|t|}{2} + 1 = b$.

Next, $\frac{|t|}{2} + (2p + 1) \leq \frac{|t|}{2} + \frac{|t|}{2} = |t|$ implies $b = \frac{|t|}{2} + 1 \leq |t| - 2p$.

Since t must contain a 1 in positions $\{|t| - p - p - p + 1, \ldots, |t| - p - p\}$, t contains a 1 in position b, i.e., the second half of t starts with a 1.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

- Let s be $0^p1^p0^p1^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. The substring vxy is entirely within the first or second half, e.g.,
     $$\overbrace{0\cdots\cdots0}^{p}\overbrace{1\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$$
     with vxy the part $\underbrace{\cdots 0\, 1 \cdots}_{vxy}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

- Let s be $0^p1^p0^p1^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \leq p$):
  1. The substring vxy is entirely within the first or second half, e.g.,
     $\overbrace{0\cdots\cdots0}^{p}\overbrace{1\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$, with $\underbrace{\cdots0\,1\cdots}_{vxy}$ spanning the end of the first block and the start of the second, in which case, uv^2xy^2z will
     move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).
  2. The substring vxy straddles the midpoint of s, i.e.,
     $0\cdots01\cdot\underbrace{\cdots10\cdots}_{vxy}\cdot01\cdots1$, in which case, uv^0xy^0z will have the
     form $0^p1^i0^j1^p$ where both i and j cannot be p, i.e., is not of the form ww.

Example Non-Context-Free Languages (cont.)

$D = \{ww \mid w \in \{0,1\}^*\}$ is not context-free.

Proof: Suppose that D is a CFL. Apply the pumping lemma for CFLs:

- Let s be $0^p1^p0^p1^p$.
- Cases of dividing s as $uvxyz$ (where $|vy| > 0$ and $|vxy| \le p$):
  1. The substring vxy is entirely within the first or second half, e.g.,
     $\overbrace{0\cdots\cdots0}^{p}\overbrace{1\cdots\cdots1}^{p}\overbrace{0\cdots0}^{p}\overbrace{1\cdots1}^{p}$ with $\underbrace{\cdots0\,1\cdots}_{vxy}$, in which case, uv^2xy^2z will move a 1 to the first position of the second half and so is not of the form ww (similarly for when vxy lies in the second half of s).
  2. The substring vxy straddles the midpoint of s, i.e.,
     $0\cdots01\cdot\underbrace{\cdots10\cdots}_{vxy}\cdot01\cdots1$, in which case, uv^0xy^0z will have the form $0^p1^i0^j1^p$ where both i and j cannot be p, i.e., is not of the form ww.

All cases lead to a contradiction. We conclude that D is not context free.

Regular vs. Context-Free Languages

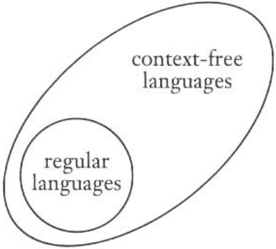

FIGURE 2.33
Relationship of the regular and context-free languages