COMPSCI 501: Formal Language Theory

Lecture 10: Pumping Lemma for CFL

Marius Minea
marius@cs.umass.edu

University of Massachusetts Amherst

13 February 2019

Recall: Pumping Lemma for Regular Languages

If A is a regular language, there is a number p (*pumping length*) so that any string in A of length at least p can be divided into three pieces, $s = xyz$, with the conditions

1. $xy^iz \in A$ for any $i \geq 0$
2. $|y| > 0$
3. $|xy| \leq p$

(1) tells us string y can be repeated (*pumped*) any number of times
(2) y is nonempty (else trivially true)
(3) says y found "early enough" (up to length p)

One-way Implication:

- **if** a language is regular, it can be pumped
- **but** if language can be pumped, it may or may not be regular

Pumping Lemma for Context-Free Languages

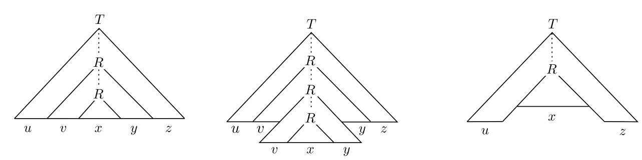
If A is a context-free language, there is a number p (pumping length) so that any string in A of length at least p can be divided into five pieces, $s = uvxyz$, with the conditions
1. $uv^ixy^iz \in A$ for any $i \geq 0$
2. $|vy| > 0$
3. $|vxy| \leq p$

(1): we can pump second and fourth string any number of times
(2): at least one of the two pumped strings is nonempty
(3): the three middle pieces have length at most p

Again, this is an *implication*, not an *equivalence*
(there are pumpable languages that are not context free)

CFL Pumping Lemma: Intuition

Work with **parse trees** of grammar



Find a nonterminal that repeats along a path, and graft trees
pump up: larger tree replaces smaller
pump down: smaller tree replaces larger

Proof: Choosing pumping length

Let V be the set of variables in the grammar
Let b be the max. number of symbols on a RHS
$\Rightarrow$ tree nodes have at most b children
$\Rightarrow$ length of sentence is $\leq b^h$ (h = tree height)

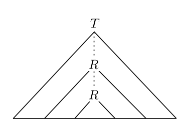
Choose $p = b^{|V|+1} > b^{|V|} \Rightarrow$ parse tree height $\geq |V| + 1$

Why not $b^{|V|} + 1$? Will see for condition (3).

Proof: Pumping is possible

Let s with $|s| \geq p$. Choose *smallest* parse tree for s.

Some path from root has length $\geq |V| + 1 \Rightarrow$
some *variable* (not terminal node) repeats.
Consider R that repeats among lowest $|V| + 1$ variables

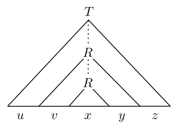


$R \overset{*}{\Rightarrow} vRy \overset{*}{\Rightarrow} vxy$
Lower R generates x
Upper R generates vxy
pump up: replace lower R with upper R
from $uvxyz$ to uv^2xy^2z to ...
pump down: replace upper R with lower R
from $uvxyz$ to uxz

Proof: Conditions 2 and 3

(2) Could it be that v and y are both empty?

If so, top R also generates $x = \varepsilon x \varepsilon$
Replacing with lower R gives smaller parsing tree for s
$\Rightarrow$ contradicts minimal choice



(3) Do we have $|vxy| \leq p$?

We chose repeated node R among lowest $|V| + 1$ variables
parse tree at top R has height $\leq |V| + 1$
generated string has $\leq b^{|V|+1} = p$ symbols

Pumping Lemma Examples: $a^n b^n c^n$

$\{a^n b^n c^n | n \geq 0\}$ is not context free

Usual string choice: $a^p b^p c^p$

- Case 1: both v and y contain at most one type of symbol

repeated symbols: at least one ($|vy| > 0$), at most two
$\Rightarrow$ won't have all three at same number

- Case 2: v or y contain more than one type of symbol pause

will have symbols out of order (ba, cb, or ca)

Pumping Lemma Examples: $a^i b^j c^k$ increasing counts

$\{a^i b^j c^k | 0 \leq i \leq j \leq k\}$ is not context free

Same string choice: $a^p b^p c^p$. Same cases, 1 more tricky

- Case 1: both v and y contain at most one type of symbol

Consider symbol that does *not* appear in either.

- a does not appear
  Pump down: uv^0xy^0z: same #a, smaller #b or #c, $\notin L$
- b does not appear
  If a appears, pump up: uv^2xy^2z has more a's than b's.
  If c appears, pump up down: uv^0xy^0z has more b's than c's.
- c does not appear
  Pump up: uv^2xy^2z has more a's or b's than c's

- Case 2: v or y contain more than one type of symbol pause

will have symbols out of order (ba, cb, or ca)

Pumping Lemma Examples: ww

$\{ww | w \in \{0,1\}^*\}$ is not context free

Usual choice $0^p 1 0^p 1$ does not work. could have $v = y = 0$.

Try $0^p 1^p 0^p 1^p$.

If vxy in first half, pumping up "lengthens" first part, so second half now starts with 1, wrong.
Symmetric for second half.

If vxy straddles center, and is $\leq p$, by pumping down we get $0^p 1^i 0^j 1^p$, with $i < p$ or $j < p$, also not ww

Wrapping up: Closure of CFL under Language Operations

- Concatenation $L = L_1 L_2 \quad S \to S_1 S_2$
- Union $L = L_1 \cup L_2 \quad S \to S_1 | S_2$
- Kleene Star $L = L_1^* \quad S \to \varepsilon | S_1 S$
- Homomorphism: replace symbols by strings
- Inverse homomorphism: replace string by symbol

Not closed under:

- Intersection (homework)
- Difference (would imply intersection): $A \cap B = A \setminus (A \setminus B)$
- Complement (with union, would imply intersection):
  $A \cap B = \overline{\overline{A} \cup \overline{B}}$

Deterministic PDA and CFL

DPDA: make NFA control deterministic instead

not the same as DFA (ε transitions for stack and input), but deterministic:

Transition function $\delta : Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \to Q \times \Gamma_\varepsilon \cup \{\emptyset\}$

From every state, can have
ε-**input moves** or consume input, not both
ε-**stack moves** or use stack, not both

Formally, for every $q \in Q, a \in \Sigma, x \in \Gamma$, exactly one of $\delta(q, a, x), \delta(q, a, \varepsilon), \delta(q, \varepsilon, x)$ and $\delta(q, \varepsilon, \varepsilon)$ is not $\emptyset$.

Accept if state accepting after end of input
Reject if not, or fails to read entire input
(incl. block on pop from empty stack, or infinite ε-loop)

Some DCFLs are not CFLs

DCFLs = languages recognized by DPDAs.

Examples: $\{0^n1^n | n \geq 0\}$ is DCFL.

$\{a^ib^jc^k | i = j \text{ or } i = k\}$ and $\{ww^r | w \in \{0,1\}^*\}$ are not DCFL.

Theorem: The class of DCFLs is closed under complementation.

Not quite as simple as swapping accept and non-accept states.

Corollary: A language whose complement isn't a CFL is not DCFL.

Example: $\{a^ib^jc^k | i \neq j \text{ or } j \neq k\}$ is not DCFL.

If it were, then $\bar{A} \cap a^*b^*c^* = \{a^nb^nc^n\}$ would be context free.