The Pumping Lemma for CFL’s

Chomsky Normal Form

Theorem: If L is a CFL, then L − {ε} has a CFG in CNF.

Proof Sketch of CNF Theorem

Step 1: remove A -> ε productions by providing alternatives when A on RHS.
Step 2: Use variable for each non-terminal instead of the non-terminal on RHS.
Step 3: Unroll long RHS’s: A -> BCDE becomes A -> BA₁; A₁ -> CA₂; A₂ -> DE;
A₁, A₂ new variables with only one production
Step 4: Substitute for A -> B unit productions

Pumping Lemma Intuition

Recall the pumping lemma for regular languages.
It told us that if there was a string long enough to cause a cycle in the DFA for
the language, then we could “pump” the cycle and discover an infinite
sequence of strings that had to be in the language.

Intuition – (2)

For CFL’s the situation is a little more complicated.
We can always find two pieces of any sufficiently long string to “pump” in
tandem.
That is: if we repeat each of the two pieces the same number of times, we get another
string in the language.

Statement of the CFL Pumping Lemma

For every context-free language L
There is an integer n, such that
For every string z in L of length ≥ n
There exists z = uvwx such that:
1. |vwx| ≤ n.
2. |vx| > 0.
3. For all i ≥ 0, uvᵢwxᵢy is in L.
Proof idea for Pumping Lemma

- Start with a CNF grammar for $L - \{\varepsilon\}$.
- Let the grammar have $m$ variables.
- Pick $n = 2^m$.
- Let $|z| \geq n$.
- We claim ("Lemma 1") that a parse tree with yield of length $z$ must have a path of length $m+2$ or more.

Proof of Lemma 1

- If all paths in the parse tree of a CNF grammar are of length $\leq m+1$, then the longest yield has length $2^{m-1}$, as in:

Back to the Proof of the Pumping Lemma

- Now we know that the parse tree for $z$ has a path with at least $m+1$ variables.
- Consider some longest path.
- There are only $m$ different variables, so among the lowest $m+1$ we can find two nodes with the same label, say $A$.
- The parse tree thus looks like:

Parse Tree in the Pumping-Lemma Proof

- Can't both be $\varepsilon$.
- $\leq 2^m = n$ because a longest path chosen

Pump Zero Times

Pump Twice
Using the Pumping Lemma

- Non-CFL’s typically involve trying to match two pairs of counts or match two strings.
- **Example:** The text uses the pumping lemma to show that \( \{ww \mid w \in \{0+1\}^*\} \) is not a CFL.

Using the Pumping Lemma – (2)

- \( \{0^i10^i \mid i \geq 1\} \) is a CFL.
  - We can match one pair of counts.
- But L = \( \{0^i10^i \mid i > 1\} \) is not.
  - We can’t match two pairs, or three counts as a group.
- **Proof** using the pumping lemma.
- Suppose L were a CFL.
- Let n be L’s pumping-lemma constant.

Using the Pumping Lemma – (3)

- Consider z = 0^n10^n10^n.
- We can write z = uvwxy, where \(|vwx| \leq n\), and \(|vx| > 1\).
- **Case 1:** vx has no 0’s.
  - Then at least one of them is a 1, and uwy has at most one 1, which no string in L does.

Using the Pumping Lemma – (4)

- Still considering z = 0^n10^n10^n.
- **Case 2:** vx has at least one 0.
  - vwx is too short (length \( \leq n \)) to extend to all three blocks of 0’s in 0^n10^n10^n.
  - Thus, uwy has at least one block of n 0’s, and at least one block with fewer than n 0’s.
  - Thus, uwy is not in L.