Theory of Computation

Pumping Lemma for CFL

Yang, Ning
(杨 宁)

School of Computer Science

2023.04

Content

- Statement
- Applications

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 $\geq$ n

There exists z = uvwxy such that:

1. $|vwx| \leq n$.
2. $|vx| > 0$.
3. For all $i \geq 0$, uv^iwx^iy is in L.

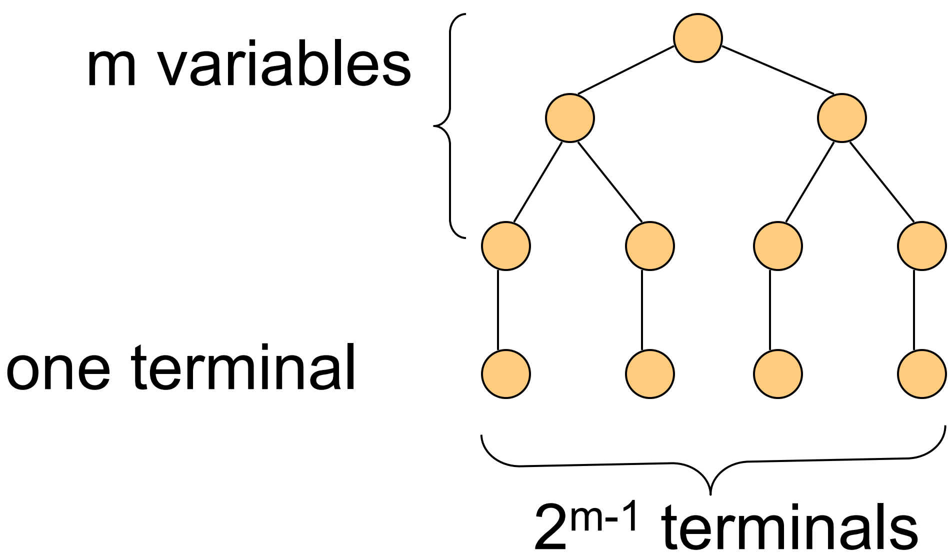
Proof of Pumping Lemma

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

Theorem 7.17: Suppose we have a parse tree according to a Chomsky-Normal-Form grammar $G = (V, T, P, S)$, and suppose that the yield of the tree is a terminal string w. If the length of the longest path is n, then $|w| \leq 2^{n-1}$.

Proof of Pumping Lemma 1

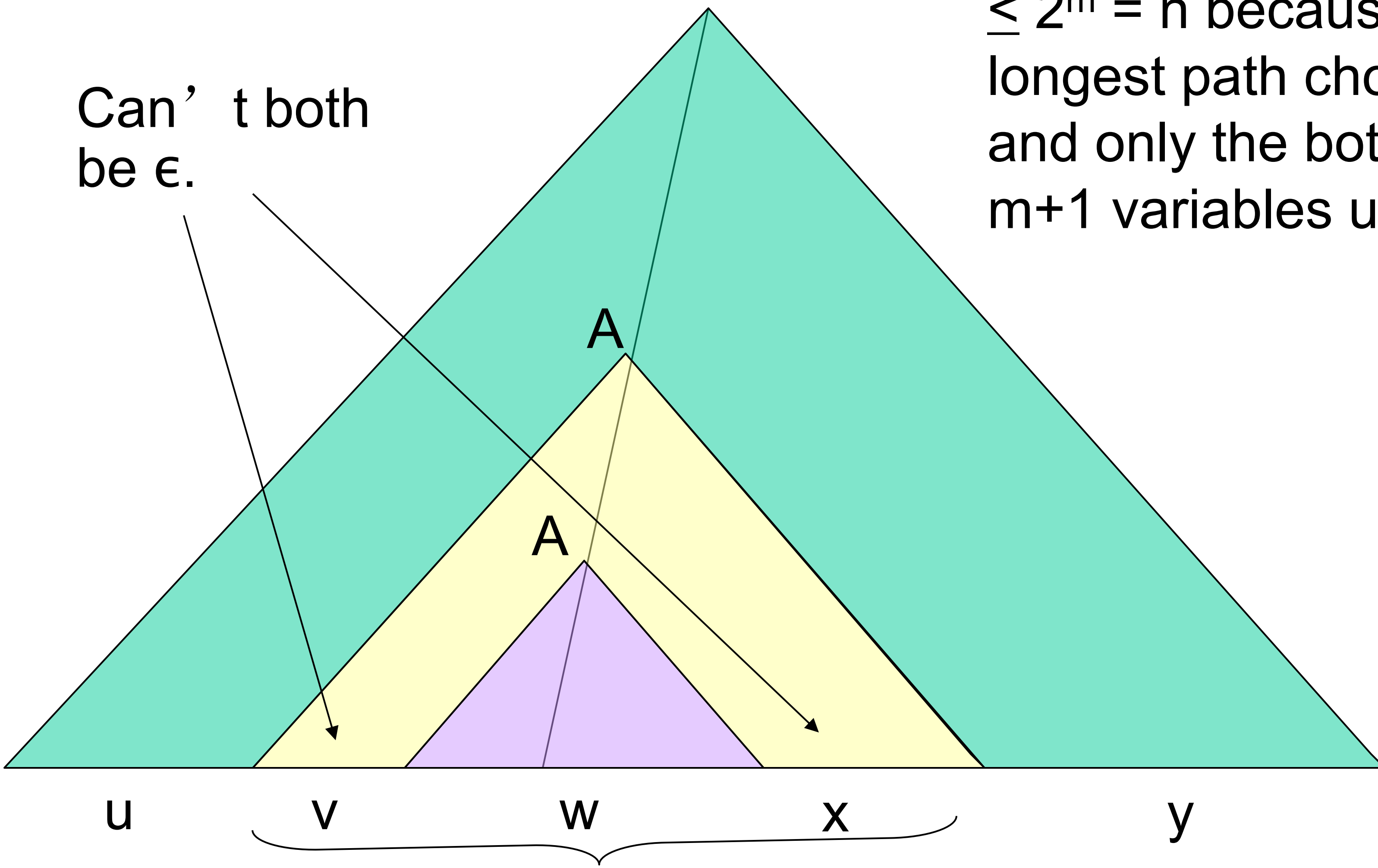
- If all paths in the parse tree of a CNF grammar are of length $\leq$ m, then the longest yield has length 2^{m-1}, as in:



Back to the Proof of Pumping Lemma

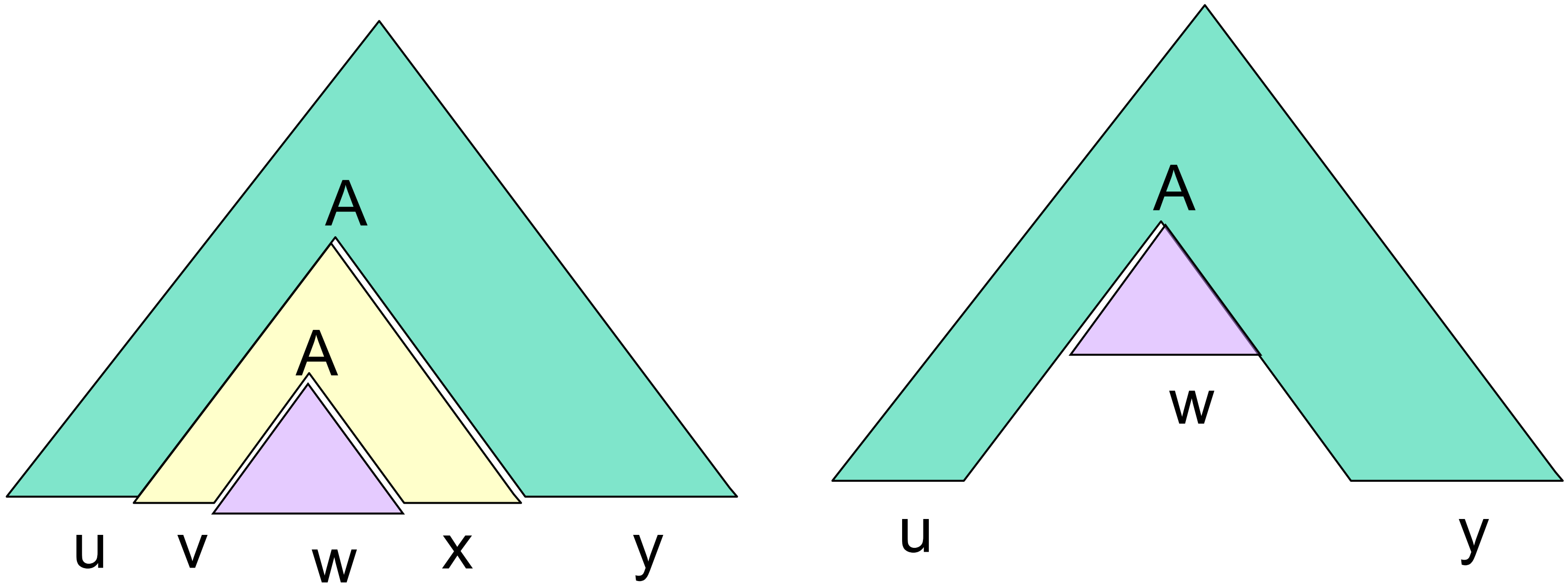
- Now we know that the parse tree for z has a path with at least m+1 variables.
- Consider some longest path.
- There are only m different variables, so among the lowest m+1 we can find two nodes with the same label, say A.
- The parse tree thus looks like:

Parse Tree in the Pumping Lemma Proof

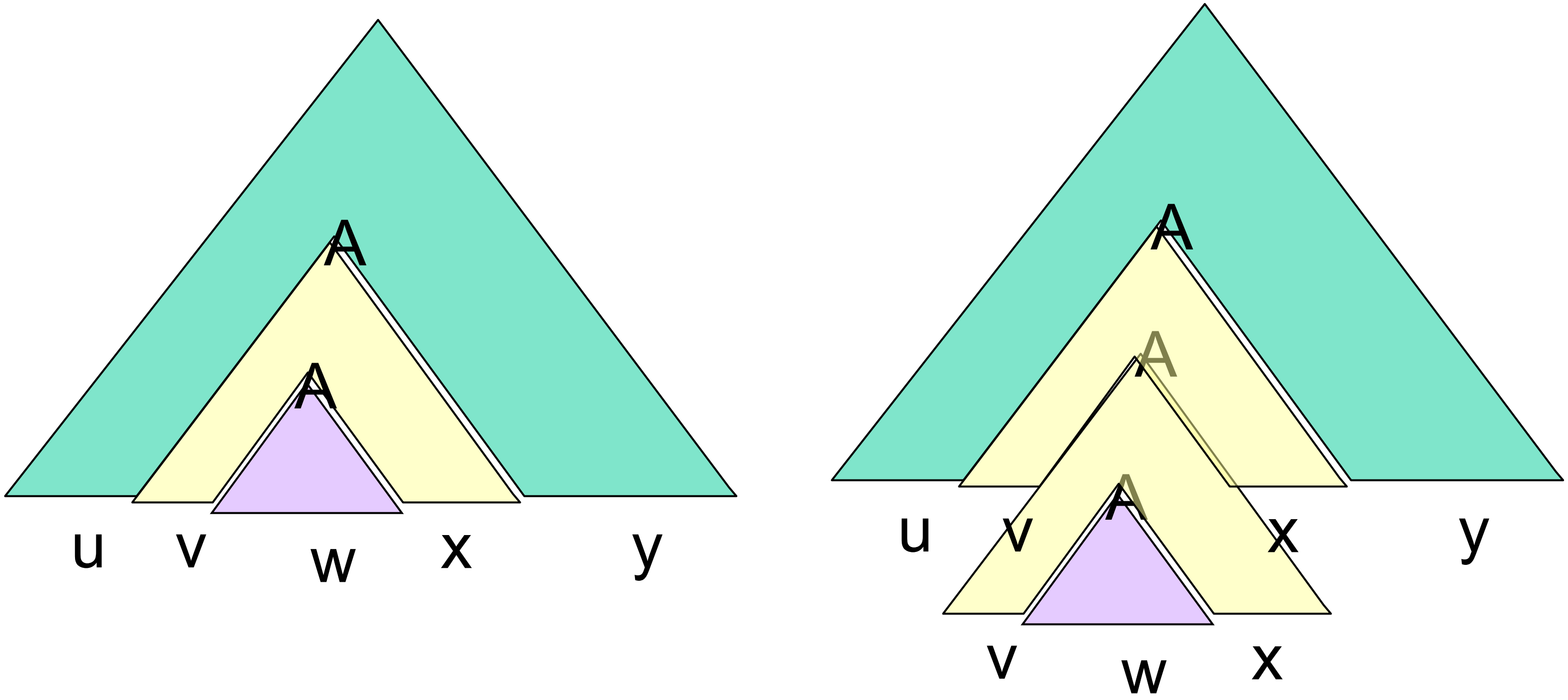


$\leq 2^m = n$ because a longest path chosen and only the bottom m+1 variables used.

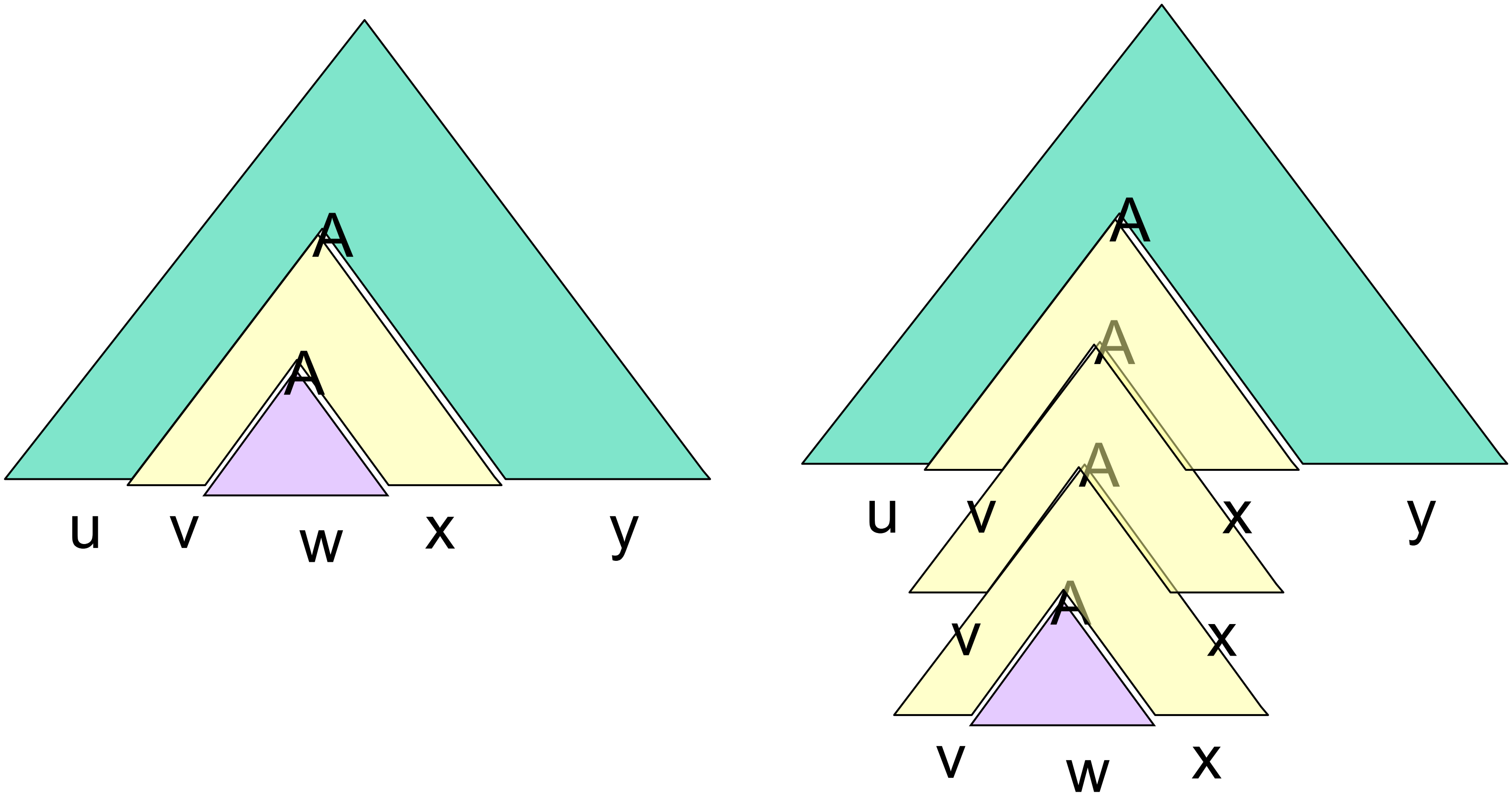
Pump Zero Times



Pump Twice



Pump Twice Etc.



Content

- Statement
- Applications

Using Pumping Lemma

- $\{0^i10^i \mid i \geq 1\}$ is a CFL.
  - We can match one pair of counts.
- But $L = \{0^i10^i10^i \mid i \geq 1\}$ 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 Pumping Lemma – (2)

- Consider $z = 0^n10^n10^n$.
- We can write $z = uvwxy$, where $|vwx| \leq n$, and $|vx| \geq 1$.
- 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 Pumping Lemma – (2)

- 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.

You Are Welcome

Any Question?