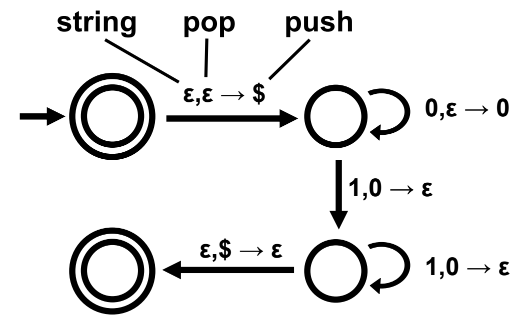
CSci 4011

INHERENT LIMITATIONS OF COMPUTER PROGRAMS



CONTEXT-FREE GRAMMARS

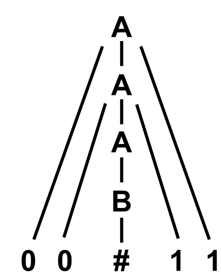
$$A \rightarrow 0A1$$
$$A \rightarrow B$$
$$B \rightarrow \#$$

$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$

uVw yields uvw if $(V \rightarrow v) \in R$.

A derives 00#11 in 4 steps.

PARSE TREES



$A \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 00B11 \Rightarrow 00\#11$

<EXPR> → <EXPR> + <EXPR>

<EXPR> → <EXPR> x <EXPR>

<EXPR> → (<EXPR>)

<EXPR> → a

Build a parse tree for a + a x a

Definition: a string is derived ambiguously in a context-free grammar if it has two or more different parse trees

Definition: a grammar is ambiguous if it generates some string ambiguously

⟨stmt⟩ → ⟨assign⟩ | ⟨if-then⟩ | ⟨if-then-else⟩

⟨if-then⟩ → if ⟨condition⟩ then ⟨stmt⟩

⟨if-then-else⟩ → if ⟨condition⟩ then ⟨stmt⟩ else ⟨stmt⟩

⟨assign⟩ → a := 1 | a := 2

⟨condition⟩ → a=1 | a=2

if a=1 then if a=2 then a := 1 else a := 2

THE CHOMSKY NORMAL FORM

A context-free grammar is in Chomsky normal form if every rule is of the form:

A → BC	B and C are not start variable
A → a	a is a terminal
S → ε	S is the start variable

S → aSb

S → ε

R → AT | AB | ε

T → SB

S → AT | AB

A → a

B → b

Theorem 2.9: Any context-free language can be generated by a context-free grammar in Chomsky normal form

"We can transform any CFG into Chomsky normal form"

Convert the following into Chomsky normal form:

$A \to BAB \mid B \mid \varepsilon$
$B \to 00 \mid \varepsilon$

$S \to SaSb \mid \varepsilon$

---

1. Add a new start variable S_0 and rule $S_0 \to S$
2. Remove all $A \to \varepsilon$ rules
3. Remove unit rules $A \to B$
4. Convert all remaining rules into the proper form

THE CHOMSKY NORMAL FORM

A context-free grammar is in Chomsky normal form if every rule is of the form:

$A \to BC$	B and C are not start variable
$A \to a$	a is a terminal
$S \to \varepsilon$	S is the start variable

Any variable A that is not the start variable can only generate strings of length > 0

Theorem: If G is in CNF, $w \in L(G)$ and $|w| > 0$, then any derivation of w in G has length $\leq 2|w| - 1$

Proof (by induction on $|w|$):

Base Case: If $|w| = 1$, then any derivation of w must have length 1

Inductive Step: Assume true for any string of length at most $k \geq 1$, and let $|w| = k+1$

Since $|w| > 1$, derivation starts with $S \to AB$

So $w = xy$ where $A \Rightarrow x$, $|x| > 0$ and $B \Rightarrow y$, $|y| > 0$

By the inductive hypothesis, the length of any derivation of w must be at most

$$1 + (2|x| - 1) + (2|y| - 1) = 2(|x| + |y|) - 1$$

THE CONTEXT-FREE PUMPING LEMMA

Let L be a context-free language

Then there exists P such that
For every w ∈ L with |w| ≥ P
there exist uvxyz=w, where:

1. |vy| > 0
2. |vxy| ≤ P
3. For every i ≥ 0, $uv^ixy^iz \in L$

EXAMPLES

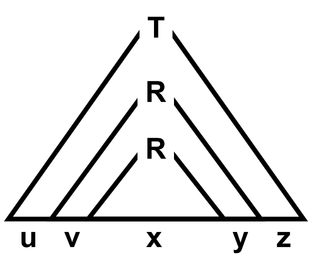
Example: L = { w ∈ {0,1}* : w = w^R}. P = 1

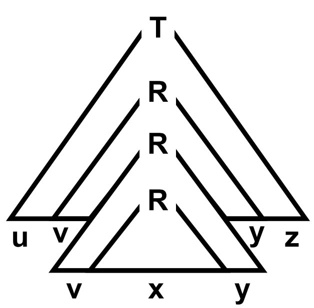
w = 0; u,v,x,y,z = ?
w = 010; u,v,x,y,z=?

Example: L = { w ∈ {a,b}* : #a > #b in w}. P = 1

w = a; u,v,x,y,z = ?
w = aab; u,v,x,y,z=?

Idea: If w is long enough, then any parse tree for w must have a path that contains a variable more than once





Formal Proof:

Let b be the maximum number of symbols on the right-hand side of a rule

If the height of a parse tree is h, the length of the string generated is at most: b^h

Let |V| be the number of variables in G

Define P = $b^{|V|+2}$

Let w be a string of length at least P

Let T be the parse tree for w with the smallest number of nodes.

T must have height at least |V|+2

The longest path in T must have ≥ |V|+1 variables
Select R to be the variable that repeats among the lowest |V|+1 variables

1. $|vy| > 0$
2. $|vxy| \le P$

Let T be the parse tree for w with the smallest number of nodes. T must have height at least |V|+2

NEGATING THE PUMPING LEMMA

USING THE PUMPING LEMMA

Prove $L = \{ww : w \in \{0,1\}^*\}$ is not context-free

Assume L is context-free.
Then there is a pumping length P.

No matter what P is, the string $w = 0^P1^P0^P1^P$ has $|w| \ge P$ and $w \in L$.

So there should be uvxyz with:
1) $|vy| > 0$, 2) $|vxy| \le P$, 3) $\forall i$, $uv^ixy^iz \in L$.

$$w = \overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}\overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}$$

USING THE PUMPING LEMMA

Prove $L = \{ww : w \in \{0,1\}^*\}$ is not context-free

Assume L is context-free.
Then there is a pumping length P.

$$w = \overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}\overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}$$

vxy

USING THE PUMPING LEMMA

Prove $L = \{w\#w^R : w \in \{0,1\}^* \text{ and } \#1s = \#0s\}$

is not context-free

Assume L is context-free.
Then there is a pumping length P.

No matter what P is, the string $w = 0^P1^P\#1^P0^P$ has $|w| \geq P$ and $w \in L$.

So there should be uvxyz with:
1) $|vy| > 0$, 2) $|vxy| \leq P$, 3) $\forall$ i, $uv^ixy^iz \in L$.

$$w = \overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}\#\overbrace{11\ldots11}^{P}\overbrace{00\ldots00}^{P}$$

USING THE PUMPING LEMMA

Prove $L = \{w\#w^R : w \in \{0,1\}^* \text{ and } \#1s = \#0s\}$

is not context-free

Assume L is context-free.
Then there is a pumping length P.

$$w = \overbrace{00\ldots00}^{P}\overbrace{11\ldots11}^{P}\#\overbrace{11\ldots11}^{P}\overbrace{00\ldots00}^{P}$$

$$w = 00\ldots0011\ldots\underbrace{11\#11}_{vxy}\ldots1100\ldots00$$