Lecture 10: Properties of Context-Free Language

Instructor: Ketan Mulmuley
Scriber: Yuan Li

February 5, 2015

1 Ogden's Lemma

Lemma 1.1. *Let L be a CFL. There exists a constant n depending only on L such that if $z \in L$, $|z| \geq n$, and if we mark any n or more positions in z as distinguished, then we can write $z = uvwxy$ such that*

- *v and x together contain at least one distinguished position.*
- *vwx has at most n distinguished position.*
- *For all $i \geq 0$, $uv^iwx^iy \in L$.*

Remark 1.2. *It implies the usual pumping lemma by marking all positions of z.*

Proof. Let G be a Chomsky normal form for $L \setminus \{\epsilon\}$ with k variables. Let $n = 2^{k+1}$. Consider any G-derivation tree of z.

Call any node of s a *branch point* if both of its sons have *distinguished* descendants.

Claim 1.3. *There exists a path with at least $k+1$ branch points on the path.*

Proof. Start at the top. If only one son of a node has distinguished descendants, then go in the direction of that son. If both sons of a node have distinguished descendants, then go in the direction of the son with more distinguished descendants. (If both sons have equal number of distinguished descendants, then go in either direction.) Therefore this path has at least $\log n = k + 1$ branch points. □

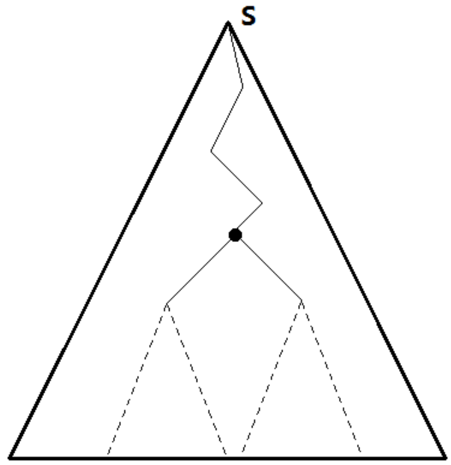

Figure 1: A path with $\geq k+1$ branch points

The rest of the proof is same as the usual pumping lemma. Look at the $k+1$ branch points on the path from the bottom. Among them, at least two have the same variable label, say A.

Then,

(1) vwx has at most $2^{k+1} = n$ distinguished descendants.

(2) Let $A \to BC$. Since A is a branch point, both B and C have distinguished descendants. It follows that either v and x has at least one distinguished descendants.

(3) Since $A \Rightarrow_G^* vAx$ and $A \Rightarrow_G^* w$, we have $A \Rightarrow_G^* v^i w x^i$, and thus $uv^i w x^i y \in L$ for all $i \geq 0$.

□

2 Closure Properties

Proposition 2.1. *Let L_1, L_2 be CFLs. Then $L_1 \cup L_2$ is also a CFL.*

Proof. Assume we have CFGs for L_1 and L_2 with start symbols S_1 and S_2 respectively. Adding production rule $S \to S_1 \mid S_2$, we will get a CFG for $L_1 \cup L_2$. □

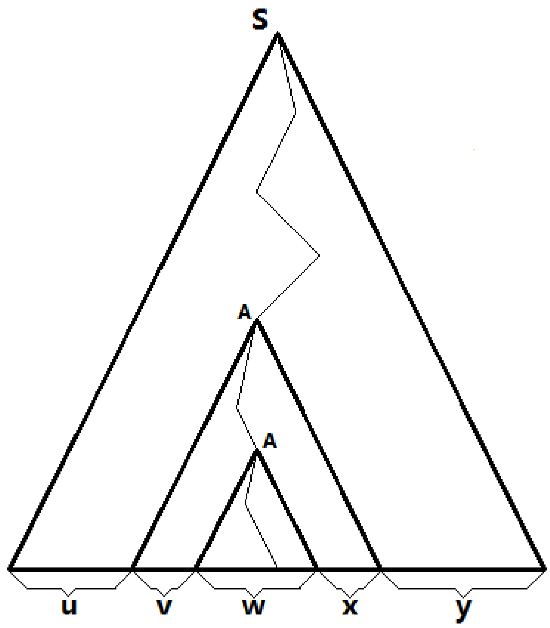

Figure 2: Proof of Ogden's lemma

Proposition 2.2. *Let L_1, L_2 be CFLs. Then $L_1L_2 = \{uv : u \in L_1, v \in L_2\}$ is also a CFL.*

Proof. Assume we have CFGs for L_1 and L_2 with start symbols S_1 and S_2 respectively. Add rule $S \to S_1S_2$, and we will get a CFG for L_1L_2. □

Proposition 2.3. *If L is a CFL, then $L^* = \cup_i L^i$ is also a CFL, where $L_0 = \{\epsilon\}$ and $L^i = LL^{i-1}$.*

Proof. Add rule $S \to \epsilon \mid SS$, where S is the start symbol. □

If both L_1, L_2 are CFLs, $L_1 \cap L_2$ may not be a CFL. For example,

$$L_1 = \{a^ib^jc^k : i = j, \text{ and } i, j, k \geq 0\},$$

$$L_2 = \{a^ib^jc^k : j = k, \text{ and } i, j, k \geq 0\},$$

are CFLs, but

$$L_1 \cap L_2 = \{a^ib^ic^i : i \geq 0\}$$

is not.

Recall that a *substitution* $f : \Sigma \to \mathcal{P}(\Delta^*)$, that is, for each a, $f(a) \subseteq \Delta^*$ is a language.

Theorem 2.4. *If L is a CFL (over alphabet Σ), and $f : \Sigma \to \mathcal{P}(\Delta^*)$ is a substitution such that $f(a)$ is a CFL for each $a \in L$, then $f(L)$ is also a CFL, where*

$$f(L) = \{w_1 w_2 \cdots w_m : w_i \in f(a_i) \text{ and } a_1 a_2 \cdots a_m \in L\}.$$

Proof. Replace $a \in \Sigma$ in the production of the grammar for L by a new symbol S_a, and add production rule $S_a \to \{$ production rule for $L_a\}$. $\square$

Corollary 2.5. *CFLs are closed under homomorphism $f : \Sigma \to \Delta^*$.*

Proposition 2.6. *CLFs are closed uner inverse homomorphisms. That is, if $L \subseteq \Delta^*$ is a CFL and $f : \Sigma \to \Delta^*$, then $f^{-1}(L) = \{w \in \Sigma^* : f(w) \in L\}$ is also a CFL.*

Proof. Let M be a PDA for L such that $L = N(M)$. Our goal is to construct a PDA M' for $L' = f^{-1}(L)$ such that $L' = N(M')$. The idea is to that, for input x, transform it to $f(x)$, and make it as an input for M. The details are left to the readers. $\square$