CS381 Fall 2000	**Handout 24**
October 25, 2000	

1. Reading: D. Kozen *Automata and Computability*, Lecture 27
J. Hopcroft and J. Ullman *Introduction to Automata Theory, etc.*, section 6.2.

2. The main message of this lecture:

> **Closure properties of context-free languages help to build new CFLs and to prove that some languages are not context-free. CFLs contain all regular languages, are closed under unions, concatenations, asterates, intersections with regular languages, homomorphic images and inverse images. CFLs are not closed under intersections and complementations.**

Theorem 24.1. *CFLs are closed under unions.*

Proof. Let $A_1 = L(G_1)$ and $A_2 = L(G_2)$ where a CFG $G_i = (N_i, \Sigma, P_i, S_i)$, $i = 1, 2$, N_1 and N_2 are disjoint. Take $G = (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \to S_1,\ S \to S_2\}, S)$. Claim: $A_1 \cup A_2 = L(G)$. Indeed, $x \in A_i \Rightarrow S_1 \xrightarrow{*}_{G_i} x \Rightarrow S \xrightarrow{*}_G x \Rightarrow x \in L(G)$, thus $A_1 \cup A_2 \subseteq L(G)$. On the other hand, let $x \in L(G)$, i.e. $S \xrightarrow{*}_G x$. Consider the very first step in a given derivation of x: it is either $S \xrightarrow{1} S_1$ or $S \xrightarrow{1} S_2$ since no other productions in G contain S. In the former case the rest of the derivation of x is entirely in G_1, since none of the productions from G_2 applies, therefore $S \xrightarrow{*}_{G_1} x$ and $x \in A_1$. Similarly, in the latter case $x \in A_2$. In either case $x \in A_1 \cup A_2$, thus $L(G) \subseteq A_1 \cup A_2$. Q.E.D.

Theorem 24.2. *CFLs are closed under concatenations.*

Proof. Let $A = L(G_1)$ and $B = L(G_2)$ where a CFG $G_i = (N_i, \Sigma, P_i, S_i)$, $i = 1, 2$, N_1 and N_2 are disjoint. Take $G = (N_1 \cup N_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \to S_1 S_2\}, S)$. Claim: $A_1 A_2 = L(G)$. Indeed, let $x_1 \in A_1$ and $x_2 \in A_1$. Then $S_i \xrightarrow{*}_{G_i} x_i$ and $S \xrightarrow{1} S_1 S_2 \xrightarrow{*}_G x_1 S_2 \xrightarrow{*}_G x_1 x_2$. Therefore, $A_1 A_2 \subseteq L(G)$. Let now $y \in L(G)$. The very first step in this derivation is necessarily $S \xrightarrow{1} S_1 S_2$, since no other production in G contains S. It is easy to see that all productions in this derivation performed on descendants of S_i are in fact productions from G_i. Therefore, the derived string y is in fact a concatenation of $x_1 x_2$, where $S_1 \xrightarrow{*}_G x_1$ and $S_2 \xrightarrow{*}_G x_2$. All productions in the former derivation are from G_1 and all productions in the latter derivation are from G_2. Therefore, $S_1 \xrightarrow{*}_{G_1} x_1$ and $S_2 \xrightarrow{*}_{G_2} x_2$, hence $y = x_1 x_2 \in A_1 A_2$. Q.E.D.

Theorem 24.3. *CFLs are closed under asterate.*

Proof. Let $A = L(G_1)$, where $G_1 = (N_1, \Sigma, P_1, S_1)$. Then A^* is generated by $G = (N_1 \cup \{S\}, \Sigma, P_1 \cup \{S \to S_1 S \mid \epsilon\ \}, S)$. Indeed, let $x \in A^*$, then $x = y_1 y_2 \dots y_n$ for some $y_i \in A$, $i = 1, 2, \dots, n$, $n \ge 0$. Therefore $S_1 \xrightarrow{*}_{G_1} y_i$. Here is a derivation of x in G

$$S \xrightarrow{n} (S_1)^n S \xrightarrow{1} (S_1)^n \xrightarrow{*} y_1 y_2 \dots y_n.$$

Let $x \in L(G)$. By induction on derivation of x in G we prove that $x \in A^*$.

Base: $S \xrightarrow{1} x$. Since x does not contain nonterminals, $x = \epsilon \in A^*$.

Step. Let $S \xrightarrow{n+1} x$. Analyzing the first step of this derivation we conclude that $S \xrightarrow{1} S_1 S \xrightarrow{n} y_1 y$, where $S_1 \xrightarrow{k} y_1$ and $S \xrightarrow{l} y$ for some $k, l > 0$ such that $k + l = n$. By the induction hypothesis, $y_1 \in A^*$ and $y \in A^*$, therefore $y_1 y \in A^*$. Q.E.D.

Corollary 24.4. *Every regular language is CFL.*

Proof. Let $A = L(\alpha)$ for some regular expression α over $\Sigma = \{a_1, a_2, \ldots, a_n\}$. By induction on the length of α we establish that A is CFL.

Base: If $\alpha = a_i$, use the grammar $S \longrightarrow a_i$. If $\alpha = \epsilon$ then use $S \longrightarrow \epsilon$. If $\alpha = \emptyset$, then use $S \longrightarrow S$ (the set of derivable terminal strings there is $\emptyset$).

Step. For '+' use 24.1, for concatenation use 24.2, for $*$ use 24.3. Q.E.D.

Theorem 24.5. *CFLs are closed under homomorphisms.*

Proof. Let $A = L(G)$ for some CFG $G = (N, \Sigma, P, S)$, $\Sigma = \{a_1, a_2, \ldots, a_n\}$, and let $h(a_i) = \alpha_i$, $i = 1, 2, \ldots, n$, be a homomorphism. Claim: $B = h(A)$ is also context free. Construct $G' = (N', \Sigma', P', S)$ for B as follows.

$N' = N \cup \{S_1, S_2, \ldots, S_n\}$, where S_i are new nonterminals

Σ' is the union of all symbols from α_i's

P' is the union of 1) the results of substituting S_i for each instance of a_i in every production from P, 2) $S_i \longrightarrow \alpha_i$, $i = 1, 2, \ldots, n$

For each G-derivation $S \xrightarrow{*} a_{n_1} a_{n_2} \ldots a_{n_k}$ the grammar G' derives

$$S \xrightarrow{*} S_{n_1} S_{n_2} \ldots S_{n_k} \xrightarrow{*} \alpha_{n_1} \alpha_{n_2} \ldots \alpha_{n_k}.$$ Q.E.D.

Theorem 24.6. *CFLs are closed under inverse homomorphic images.*

Proof. Hopcroft-Ullman, pp. 132-134.

Theorem 24.7. *CFLs are closed under intersections with regular sets.*

Proof. Let $A = L(M)$ for an NPDA M and $R = L(N)$ for a DFA N. We build an NPDA M' for $A \cap R$ by a product construction: the states of M' are pairs of states from M and from N. The general idea of M' is to run M and N in parallel on the same input. The first component of M' simulates moves of M, including takings care of the stack changes. The second component of M' simulates moves of N without paying any attention on the stack. M' accepts x only when both M and N accept x, i.e. when $x \in A \cap R$. For the details, see Hopcroft-Ullman, pp. 134-135. This construction does not extend to the case of two NPDA's. Why? Q.E.D.

Theorem 24.8. *CFLs are NOT closed under intersections.*

Proof. $A = \{a^n b^n c^m \mid m, n \geq 0\}$ and $B = \{a^n b^m c^m \mid m, n \geq 0\}$ are both CFL, but $A \cap B == \{a^n b^n c^n \mid n \geq 0\}$ is not CFL. Q.E.D.

Theorem 24.9. *CFLs are NOT closed under complementations.*

Proof. Otherwise CFLs would be closed under intersections, since $A \cap B = \sim(\sim A \cup \sim B)$.

Homework problems. Kozen, p.335 ♯ 76a; p.336 ♯ 84fgh; ♯ 85abfm.