7 Non-context-free languages

7.1 Proving non-context-free by pumping lemma

Reading: Sipser 2.3 (pp. 123-127)
Lemma: Consider a parse tree according to a CNF grammar with a yield of $w \in \Sigma^*$. If the height of the tree is h then $|w| \leq 2^{h-1}$.
Theorem (The pumping lemma for context-free languages): Let A be a CFL. Then there exists a constant p (the pumping length) such that $\forall s \in A$ with $|s| \geq p$, we can write $s = uvxyz$ such that

1. $|vy| > 0$;
2. $|vxy| \leq p$; and
3. $\forall i \geq 0$, string $uv^ixy^iz \in A$.

Proof: Given a CFL A, there is a CNF $G = (V, \Sigma, R, S)$ such that $L(G) = A - \{\varepsilon\}$. Let $m = |V|$ and choose $p = 2^m$. Suppose $s \in L(G)$ with $|s| \geq p$. Any parse tree for s must have a height of at least $m+1$, otherwise $|s| \leq 2^{m-1} = p/2$ by Lemma. Therefore, there must be a path $A_0(S), A_1, \ldots, A_l, a$ with at least $m+1$ variables, i.e., $l \geq m$, in the parse tree for s. Since there are only m different variables in the grammar, by the pigeonhole principle, there must be at least two identical variables on the path. Choose the identical pair closest to the leaf, i.e., $A_i = A_j$ with $l - m \leq i < j \leq l$. Then it is possible to divide the parse tree such that x is the yield of the subtree rooted at A_j, vxy is the yield of the subtree rooted at A_i, and $s = uvxyz$ is the yield of the entire parse tree. Next we examine whether this partition satisfies all three conditions stated in the pumping lemma. First, since there is no unit production in CNF, v and y can not both be ε. So $|vy| > 0$. Second, since the height of the subtree rooted at A_i is at most $m+1$, its yield vxy has length at most $2^m = p$. So $|vwx| \leq p$. Third, since $A_i = A_j = T$, we have $T \overset{*}{\Rightarrow} vTy$ and $T \overset{*}{\Rightarrow} x$ from the parse tree. So for any $i \geq 0$, $T \overset{*}{\Rightarrow} v^ixy^i$, thus, $S \overset{*}{\Rightarrow} uTz \overset{*}{\Rightarrow} uv^ixy^iz$. So $uv^ixy^iz \in A$ for $i \geq 0$. This completes the proof of the pumping lemma.
How to use the pumping lemma to prove that a language is not context-free:

- Assume that A was context-free by contradiction. Then the pumping lemma applies to A. Let p be the constant in the pumping lemma.
- Select $s \in A$ with $|s| = f(p) \geq p$.
- By the pumping lemma, $s = uvxyz$ with $|vy| > 0$, $|vxy| \leq p$, and $uv^ixy^iz \in A\ \forall i \geq 0$.
- For any u, v, x, y, z such that $s = uvxyz$, $|vy| > 0$, and $|vxy| \leq p$, find $i \geq 0$ such that $uv^ixy^iz \notin A$. A contradiction!

Example (Sipser p. 126): Prove that $B = \{a^nb^nc^n | n \geq 0\}$ is not context-free.
Example (Sipser p. 127): Prove that $D = \{ww | w \in \{0,1\}^*\}$ is not context-free.
Example: Prove that $\{0^i1^j | j = i^2\}$ is not context free.

7.2 Proving non-context-free by closure properties

Note that the following closure properties of CFLs may also be used to prove that a given language is context free.

- Closed under union: If A and B are context-free, so is $A \cup B$.
- Closed under concatenation: If A and B are context-free, so is AB.
- Closed under star: If A is context-free, so is A^*.
- Closed under reverse: If A is context-free, so is A^R.
- Not closed under intersection: Consider $A = \{a^nb^nc^m\}$ and $B = \{a^mb^nc^n\}$.
- Not closed under complementation: Note that $A \cap B = \overline{\overline{A} \cup \overline{B}}$.
- Not closed under difference: Note that $\overline{A} = \Sigma^* - A$.

- Intersect with a regular language: If A is context-free and B is regular, then $A \cap B$ is context-free.
- Difference from a regular language: If A is context-free and B is regular, then $A - B$ is context-free. Note that $A - B = A \cap \overline{B}$.

Example: $A = \{w \in \{a,b,c\}^* | w$ has equal numbers of as, bs and cs$\}$ is not a CFL.