CS525 Winter 2012 \ Class Assignment #2, 1/25/2012

Aylin Caliskan

Nicholas Milkovits

Ariel Stolerman

2.26)

Let G be a CFG in Chomsky Normal Form. Following is a proof that for any $w \in L(G)$ of length $n \geq 1$ exactly $2n - 1$ steps are required for any derivation of w. We prove by induction on the length of w, n.

Since G is a CFG in Chomsky Normal Form, then any rule is of the form: $A \rightarrow BC$ (where the variables $B, C \neq S$ and S is the start symbol), $A \rightarrow a$ (where the terminal $a \neq \epsilon$) and it may contain $S \rightarrow \epsilon$.

First, for any $w \in L$ such that $|w| = 1$, it must be derived by a rule $S \rightarrow a$, where S is the start symbol and a is a terminal different than ϵ. It cannot be that $a = \epsilon$, otherwise $|w| = 0$. Also, it cannot be derived starting at a rule $S \rightarrow AB$, since neither A nor B can derive ϵ, therefore any derivation starting at $S \rightarrow AB$ derives a string with length ≥ 2. Therefore the derivation must be by the sequence of rules $< S \rightarrow a >$ (where $w = a$), which contains only one step, and satisfies the required $2 \cdot 1 - 1 = 1$ rules in any derivation of w of size 1.

Now assume that any string of size strictly less than m in L satisfies the claim, and let w be a string in L of size m. The case in which w is a single terminal is handled in the base case of the induction. Therefore for any w of size greater than 1 the first step in the derivation has to be of the form $S \rightarrow AB$, where S is the start symbol and A, B are variables. Moreover, since A, B cannot be S, they cannot derive a string of size 0 (ϵ), thus we can say that $A \rightarrow x$ and $B \rightarrow y$ where x, y are non-empty strings, and $w = xy$ (the concatenation of x and y). Denote $|x| = i$ and $|y| = j$, so $i + j = m$ (and $i, j > 0$). Therefore according to the induction assumption x is derived in $2i - 1$ steps and y is derived in $2j - 1$ steps. Since w is derived by the step $S \rightarrow AB$ followed by the derivations of x and y (in some order), the total derivation steps for w are:

$$1 + (2i - 1) + (2j - 1) = 2(i + j) - 1 = 2m - 1$$

Thus proving the claim.

2.31

Let $B = \{w \in \{0,1\}^* \mid w\ is\ a\ palindrome\ with\ \#0's = \#1's\}$. Following is a proof that B is not a CFL:

Assume B is a CFL, then it satisfies the conditions of the pumping lemma. Let p be the pumping length of B and let $w = 0^p 1^{2p} 0^p$. Clearly $w \in B$ since it is a palindrome ($0^p 1^p$ reflects to $1^p 0^p$) and has the same number of 0's and 1's ($2p$). But there exists no partition of $w = uvxyz$ that satisfies the pumping lemma conditions and for which $uv^2xy^2z \in B$: since $|vxy| \leq p$ it may be constructed of only 0's, 0's followed by 1's or 1's followed by 0's. The cases:

- If vxy consists only of 0's, since at least one of v, y has length at least 1, then the pumped string will have more 0's on the left half than on the right half, and will have more 0's than 1's, thus the pumped string will not be in B. This case is true for vxy consisting of 0's only from the right half or only from the left half.

- If vxy consists only of 1's, similarly to the previous case, the string will have more 1's than 0's thus the pumped string will not be in B (although it might still be a palindrome).
- If vxy consists partially of 0's and partially of 1's, i.e. $vxy = 0^i1^j$ where $1 \leq i + j \leq p$, since at least one of v, y has length at least 1, the pumping will result with:
  - If one of v, y spans over both 0's and 1's, the pumping will create a non-palindrome string, since it will add some pattern of $0^i1^j0^i1^j$ to the left hand side of the string, which will not be matched on the right hand side.
  - If v is only 0's and y is only 1's, the pumping will again create a string not in B, as either the left hand side will not be matched on the right, or it will violate the equality between the number of 0's and 1's.

Therefore in any case the pumped string will not be in B, in contradiction to the assumption. Therefore B is not a CFL.

2.35

Let G be a CFG in Chomsky normal form that contains b variables and let s be some string generated by G with at least 2^b steps. Following is a proof that $L(G)$ is infinite:

Following the proof of the pumping lemma for CFLs in the book, any binary tree of height b has at most 2^b leaves (where the root is of height 0), and the number of internal nodes is at most $2^b - 1$ (when the number of leaves is indeed 2^b). Therefore if the number of internal nodes is at least 2^b, the height of the tree will be at least $b + 1$.

Since the grammar is in Chomsky normal form, the maximum number of symbols on the right hand side of any rule is 2 (2 variables), thus any node in a parse tree of the grammar will have at most 2 children. If there exists a string $s \in L(G)$ such that it is derived with at least 2^b steps, it must have at least 2^b internal nodes, therefore the height of the parse tree will be at least $b + 1$. Continuing as in the pumping lemma proof, this means any parse tree of s would have a path of length $b + 1$, i.e. constructed of at least $b + 2$ symbols where the last one (leaf) is a terminal, and the first $b + 1$ (or more) are variables. Since there are only b variables, following the pigeon-hole principle, at least one variable appears along the path at least twice. Let R be a repeating variable among the lowest $b + 1$ variables, then we can replace the lower appearance with the higher one as many times as we like, pumping s forever, where each pumped string is in $L(G)$ (since we are getting a legal parse tree), thus proving $L(G)$ is infinite.