CS525 Winter 2012 \ Class Assignment #2 Preparation

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:

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

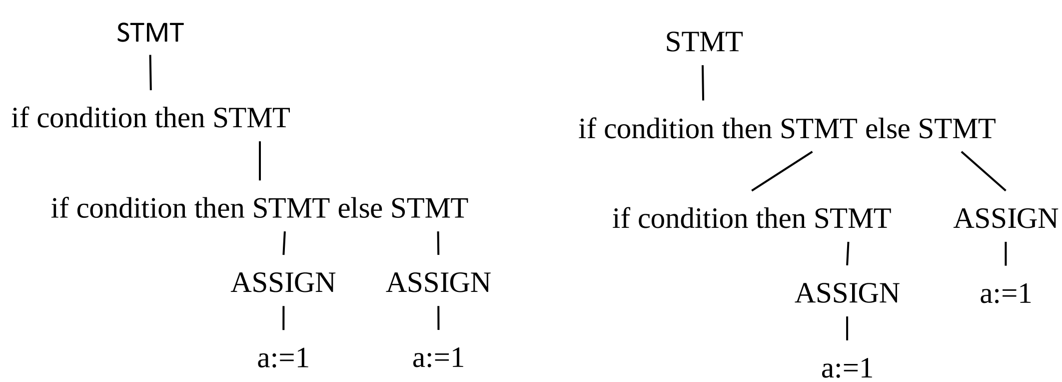
Thus proving the claim.

2.27)

a.

Following is an example for an ambiguous string in $L(G)$ for the given grammar G with 2 different derivations:

$s =$ if condition then if condition then a:=1 else a:=1



b.

Following is an unambiguous grammar for the same language:

<STMT> → <ASSIGN> | <IF-THEN> | <IF-THEN-ELSE>
<IF-THEN> → if condition then <STMT> | if condition then <IF-THEN-ELSE> else <IF-THEN>
<IF-THEN-ELSE> → if condition then <IF-THEN-ELSE> else <IF-THEN-ELSE> | <ASSIGN>
<ASSIGN> → a:=1

2.28)

Following are unambiguous grammars for the given languages.

a.

$L = \{w \mid in\ every\ prefix\ of\ w\ \#a's \geq \#b's\}$

$S \to aBS \mid aS \mid \epsilon$

$B \to aBB \mid b$

b.

$L = \{w \mid \#a's = \#b's\ in\ w\}$

$S \to aAbS \mid bBaS \mid \epsilon$

$A \to aAb \mid aBb \mid ab$

$B \to bBa \mid bAa \mid ba$

c.

$L = \{\#a's \geq \#b's\ in\ w\}$

$S \to aAbS \mid bBaS \mid aS \mid \epsilon$

$A \to aAb \mid aBb \mid ab$

$B \to bBa \mid bAa \mid ba$

2.29

Let $L = \{a^i b^j c^k \mid i = j\ or\ j = k, where\ i, j, k \geq 0\}$. Following is a proof that L is inherently ambiguous:

Following is an example grammar for the string:

$S \to AB \mid CD$
$A \to aAb$
$B \to Bc \mid \epsilon$
$C \to Ca \mid \epsilon$
$D \to bDc$

This grammar is ambiguous, yet it does not prove that no unambiguous grammar for L exists. However, the only way to ensure for the strings where $i = j$ that indeed $\#a's = \#b's$ is with a rule like $A \to aAb$. For instance, $a^n b^n c^m \in L$ must be derived from any grammar of L, for any m (representing the number of $c's$), thus it also must have rules that can derive any number of $c's$ like $B \to Bc \mid \epsilon$. Similarly for strings where $j = k$ we need a rule like $D \to bDc$ to ensure $\#b's = \#c's$. For instance, $a^m b^n c^n \in L$ must be derived from any grammar of L, for any m (representing the number of $a's$), thus it also must have rules that can derive any number of $a's$ like $C \rightarrow Ca \mid \epsilon$.

Therefore, any grammar for L must be able to derive strings of the form $a^n b^n c^n \in L$ in at least two ways: once using n times the derivation of the same number of $a's$ and $b's$ followed by n times the derivation of $c's$, and once using n times the derivation of the same number of $b's$ and $c's$ followed by n times the derivation of $a's$. Therefore any grammar for L is ambiguous, thus L is inherently ambiguous.

2.30

Following are proofs using the pumping lemma that L is not a CFL.

a.

Let $L = \{0^n 1^n 0^n 1^n \mid n \geq 0\}$.

Assume that L is a CFL, and let $w = 0^p 1^p 0^p 1^p$ where p is the pumping length of L. Clearly $w \in L$, but any partition of w into $uvxyz$ where $|vy| > 0$ and $|vxy| \leq p$ satisfies that $uv^2xy^2z \notin L$: if we denote the first 0^p as a, first 1^p as b, second 0^p as c and second 1^p as d then since $|vxy| \leq p$ the whole substring vxy crosses at most only two consecutive sectors, i.e. it is all part of $a, a+b, b, b+c, c, c+d$. In all of those cases, since $|vy| > 0$ then either $|v| \geq 1$, $|y| \geq 1$ or both. Therefore it cannot be that the pumping will change all sectors a, b, c, d (at most two will change) thus $uv^2xy^2z \notin L$, in contradiction to the assumption. Therefore L is not a CFL.

d.

Let $L = \{t_1 \# t_2 \# \dots \# t_k \mid k \geq 2, \forall i: t_i \in \{a,b\}^*, t_i = t_j \; for \; some \; i \neq j\}$.

Assume that L is a CFL and let $w = a^p b^p \# a^p b^p$ where p is the pumping length of L. Clearly $w \in L$, but any partition of w into $uvxyz$ where $|vy| > 0$ and $|vxy| \leq p$ satisfies that $uv^0xy^0z \notin L$:

Since this string contains only 2 t's, it is sufficient to show that the pumping will make t_1 (the substring before the #) and t_2 (the substring after the #) different.

- If vxy is contained completely in t_1 (before the #), since at least one of v, y has length ≥ 1, the pumping will alter t_1 to be different than t_2 with either different number of a's or b's (or both), thus $t_1 \neq t_2$.
- If vxy is partially contained in t_1 and partially in t_2:
  - If either v or y contain the #, pumping (down) will eliminate it and the pumped string will contain only a single block of $a's$ and $b's$ such that the k in the language description will be equal to 1, thus the pumped string will not be in L.
  - If neither v nor y contain the #, since at least one of v, y has length at least 1, the pumping will alter the number of a's in t_2 or the number of $b's$ in t_1 such that $t_1 \neq t_2$.
- If vxy is contained completely in t_2, it is similar to the first case.

Therefore no partition of $a^p b^p \# a^p b^p$ into $uvxyz$ by the conditions of the pumping lemma can derive uv^0xy^0z that will also be in L in contradiction to the assumption. Thus L is not a CFL.

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 $|vxy| \le p$ 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. $vxy = 0^i 1^j$ where $1 \le i + j \le p$, 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 $0^i 1^j 0^i 1^j$ 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.

2.32

Let $C = \{w \in \{1,2,3,4\}^* \mid \#1's = \#2's\ and\ \#3's = \#4's\}$. Following is a proof that C is not a CFL:

Assume C is a CFL, then it satisfies the conditions of the pumping lemma. Let p be the pumping length of C and let $w = 1^p 3^p 2^p 4^p$. Clearly $w \in C$. But there exists no partition of w into $uvxyz$ such that satisfies the pumping lemma conditions and for which $w' = uv^2xy^2z \in C$: since $|vxy| \le p$ it may be constructed of only 1's, 1's and 3's, only 3's, 3's and 2's, only 2's, 2's and 4's or only 4's. The cases:

- If vxy contains only 1's, since at least one of v, y has length at least 1, then in w' there will be more 1's than 2's, thus $w' \notin C$. A similar violation happens when vxy is consisted of only 2's, 3's or 4's.
- If vxy contains both 1's and 3's, since at least one of v, y has length at least 1:
  - If one of v, y contains both 1's and 3's, then w' will contain more 1's than 2's and more 3's than 4's, therefore $w' \notin C$.
  - If one of v, y contains only 1's or only 3's, then w' will contain more 1's than 2's or more 3's than 4's, therefore $w' \notin C$.

  A similar violation happens when vxy is consisted of only 3's and 2's or only 2's and 4's.

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

2.33

Let $F = \{a^i b^j \mid i = kj\ for\ some\ positive\ integer\ k\}$ (fixed phrasing from the textbook errata webpage). Following is a proof that C is not a CFL:

Assume F is a CFL, then it satisfies the conditions of the pumping lemma. Let p be the pumping length of C and let $w = a^{q^2} b^q$ where $p < q$. Clearly $w \in F$, but for any partition of w into $uvxyz$ such that $|vy| > 0$ and $|vxy| \leq p$ there exists a pumping of w to $w' = uv^i xy^i z$ such that $w' \notin F$:

- If vxy is consisted only of a's, since at least one of v, y has length at least 1 and $|vy| \leq p$, choosing $i = 0$ will remove between $1 \leq k \leq p$ a's such that $w' = a^{q^2-k} b^q$ will have between $q(q-1) = q^2 - q < q^2 - p \leq q^2 - k \leq q^2 - 1 < q^2$, therefore $q^2 - k$ cannot be a multiple of q, thus $w' \notin F$.
- If vxy is consisted only of b's, since at least one of v, y has length at least 1 and $|vy| \leq p$, choosing $i = q^3$ will in any case pump the number of b's to be more than the number of a's (the least number of b's will be in $w' = a^{q^2} b^{q-1+q^3}$), so that it cannot be an integer multiplication of the number of b's. Thus $w' \notin F$.
- If vxy is consisted of both a's and b's, since at least one of v, y has length at least 1:
  - If one of v, y contains both a's and b's, choosing $i = 2$ will make w' contain some b's before a's so w' will not be of the form $a^i b^j$, thus $w' \notin F$.
  - If v contains only a's and y contains only b's (and $|v|, |y| > 0$ otherwise it is one of the previous cases), denote $|v| = n, |y| = m$. Note that $k = \frac{\#a's}{\#b's} = \frac{q^2 - n + jn}{q - m + jm} \underset{i:=j+1}{=} \frac{q^2 + in}{q + im}$ and for $i \geq 0$ we must show there exists some i such that this k is not an integer. For $i = 0$ we get $k = \frac{q^2}{q} = q$; for $i \to \infty$ we get $\frac{n}{m}$. Since $1 \leq n, m \leq p - 1 < q$, we get $\frac{1}{p-1} \leq \frac{n}{m} \leq p - 1$. In the largest range possible, k will be in $\left(\frac{n}{m}, q\right] = \left(\frac{1}{p-1}, q\right]$, which has a finite number of integers. Since all $i \geq 0$ derive distinct k's, there must be some i for which k will not be an integer, thus $w' \notin F$.

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

2.34

Let G be a grammar of the language $B = L(G)$ with the following rules:

$S \to TT \mid U$

$T \to 0T \mid T0 \mid \#$

$U \to 0U00 \mid \#$

Find the minimum pumping length p of B that will work in the pumping lemma.

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.