Chapter 16: Non-Context-Free Languages *

Peter Cappello
Department of Computer Science
University of California, Santa Barbara
Santa Barbara, CA 93106
cappello@cs.ucsb.edu

• Please read the corresponding chapter before attending this lecture.

• These notes are not intended to be complete. They are supplemented with figures, and material that arises during the lecture period in response to questions.

*Based on Theory of Computing, 2nd Ed., D. Cohen, John Wiley & Sons, Inc.
**Self-Embeddedness**

- We call a production **live** when it is of the form
  \[ \text{Nonterminal} \rightarrow \text{Nonterminal Nonterminal} \]
- We call a production **dead** when it is of the form
  \[ \text{Nonterminal} \rightarrow \text{terminal} \]

**Theorem**

Let \( G \) be a CFG in CNF. Let \( S = \{ w \mid \text{Each live production of } G \text{ is used at most once to generate } w \} \). \( S \) is finite.
Proof

1. The 1st production used produces a working string of:
   • length 1, if a dead production is applied
   • length 2, if a live production is applied.

2. Every live production applied increases the length of the working string by 1. Illustrate.

3. Every dead production applied leaves the length of the working string unchanged. Illustrate.

4. The length of the working string = 1 + the number of live productions that have been applied. Illustrate for 0 productions applied to $S$; 1 production applied.

5. If the number of live productions is $p$, then $S$ contains no word of length $> p + 1$, and hence is finite.
Theorem

Let $G$ be a CFG in CNF that has $p$ live productions. Let $w \in L(G)$ be such that $|w| > 2^p$. In every parse tree for $w$, there is some nonterminal that is a descendent of itself.

Illustrate this.

Proof

1. Since $G$ is in CNF the parse tree is a binary tree.

2. Since $|w| > 2^p$, its parse tree must have a nonterminal, $N$, of depth $> p$. Illustrate this.

3. Therefore, some nonterminal on the path from the root to this non-terminal node is repeated.
Example

Consider the CFG in CNF for the non-empty palindromes over \(\{a, b\}\):

\[
S \rightarrow AX \mid BY \mid AA \mid BB \mid a \mid b
\]

\[
X \rightarrow SA
\]

\[
Y \rightarrow SB
\]

\[
A \rightarrow a
\]

\[
B \rightarrow b
\]

• Consider a derivation for the word \(aba\).

• \(S \Rightarrow AX \Rightarrow aX \Rightarrow aSA \Rightarrow abA \Rightarrow aba\).

• What does the parse tree look like?
**Definition**

For a derivation of word $w$, a nonterminal is **self-embedding** if it occurs as a descendent of itself in the parse tree corresponding to the derivation.

- In the derivation of $aba$ in the previous example, $S$ is self-embedding.

**Definition**

Let $S_1 \Rightarrow \cdots \Rightarrow S_n$. We say $S_1 \Rightarrow^* S_n$: $S_1$ **eventually derives** $S_n$.

- Using our example: $S \Rightarrow AX \Rightarrow aX \Rightarrow aSA \Rightarrow abA \Rightarrow aba$.

- We thus can say:
  - $S \Rightarrow aSa$
  - $S \Rightarrow aaSaa$
  - $S \Rightarrow a^nSa^n$
The Pumping Lemma for CFLs

Theorem

Let $G$ be a CFG in CNF with $p$ live productions. Let $w \in L(G)$ with $|w| > 2^p$. Then, $w = uvxyz$, where:

- $x \neq \Lambda$
- $v \neq \Lambda$ or $y \neq \Lambda$
- $uv^ny^nz \in L(G)$, for all $n > 0$. 


**Proof**

1. Since $|w| > 2^p$, in every parse tree for it, there is some self-embedded nonterminal, $N$.

   Illustrate the parse tree.

2. Can $v = \Lambda$ under any circumstance? Illustrate.

3. Clearly, the subtree rooted at the 1st occurrence of $P$ can be replicated at its 2nd occurrence.

4. Hence, we can pump the derivation, to obtain $uv^nxy^nz$, for any $n > 0$.

5. Can we replace the previous constraint on $n$ with $n \geq 0$?
Algebraic Proof

1. Since $|w| > 2^p$, in every parse tree for it, there is some self-embedded nonterminal, $N$.

2. Since $N$ is self-embedded:
   
   (a) $S \Rightarrow uNz$
   
   (b) $N \Rightarrow vNy$
   
   (c) $N \Rightarrow x$.

3. Therefore,
   
   $S \Rightarrow uNz \Rightarrow uvNyz \Rightarrow uv^N y^n z \Rightarrow uv^n xy^n z$.
Example

Consider the following CFG in CNF:

\[
S \rightarrow PQ \\
Q \rightarrow QS \mid b \\
P \rightarrow a
\]

- \( S \Rightarrow PQ \Rightarrow aQ \Rightarrow aQS \Rightarrow abS \Rightarrow abPQ \Rightarrow abaQ \Rightarrow abab \)
- Draw the parse tree for this derivation.
- There are 3 instances of self-embedding. (Is \( P \) self-embedded?)
- If we use \( S \) as the self-embedded nonterminal, \( u = y = z = \Lambda \) and \( v = x = ab \). Isolate the embedded trees.
• If we use $Q$ as the self-embedded nonterminal, where the lower right $Q$ is the 2nd occurrence, we get a $u = a$, $v = ba$, $x = b$ and $y = z = \Lambda$. Isolate the embedded trees.

• If we use $Q$ as the self-embedded nonterminal, where the lower left $Q$ is the 2nd occurrence, $u = a$, $v = \Lambda$, $x = b$, $y = ab$, $z = \Lambda$. Isolate the embedded trees.
Theorem

Let $G$ be a CFG in CNF with $p$ live productions. Let $w \in L(G)$ with $|w| > 2^p$. Then, $w = uvxyz$, where:

- $x \neq \Lambda$
- $v \neq \Lambda$ or $y \neq \Lambda$
- $|vxy| \leq 2^p$
- $uv^nv^nz \in L(G)$, for all $n > 0$

Proof

1. There is a leaf in the parse tree along a path with a self-embedded nonterminal $\leq p + 1$ nodes up from the leaf. Illustrate this.

2. The subtree rooted at that nonterminal is of height $p+1$ (including the terminal row), and has at most $2^p$ leaves (i.e., terminals). Illustrate.
**APPLICATION**

$L = \{a^n b^n a^n \mid n > 0\}$ IS NOT A CFL

1. Assume $L = L(G)$, for some CFG $G$ in CNF with $p$ live productions.
2. Pick $n = 2^p$.
3. $a^n b^n a^n = uvxyz$, where $|vxy| \leq n$.
4. The $vxy$ part of this word is either:
   - entirely within a region of $a$’s or entirely within a region of $b$’s
   - spans a region of $a$’s and $b$’s, but does not span all 3 regions.
5. In the 1st case, if it is pumped, the resulting word is of the form $a^i b^j a^k$, where not all $i$, $j$, and $k$ are equal. **Illustrate.**
6. In the 2nd case, if it is pumped, the resulting word **may** still be of the
form $a^* b^* a^*$, but the block that is not part of $vxy$ will not be pumped. Consequently, $a^i b^j a^k$, where not all $i$, $j$, and $k$ are equal. Illustrate.
APPLICATION

\( \{a^m b^n a^m b^n \mid n > 0\} \) is not a CFL

1. An analysis similar to the previous application applies here too. Illustrate.

2. Therefore this language is not a CFL.