Pumping Lemma for CFLs

In any sufficiently long string in a CFL, it is possible to find at most two short, nearby substrings that we can "pump" i times in tandem, for any integer i, and the resulting string will still be in that language.

Pumping lemma for CFLs: Let L be a CFL. Then there exists a constant n such that if $z \in L$ with $|z| \geq n$, then we can write $z = uvwxy$, subject to the following conditions:

1. $|vwx| \leq n$.

2. $vx \neq \epsilon$.

3. For all $i \geq 0$, we have $uv^iwx^iy \in L$.

Informal Proof

If the string z is sufficiently long, then the parse tree produced by z has a variable symbol that is repeated on a path from the root to a leaf. Suppose $A_i = A_j$, such that
the overall parse tree has yield $z = uvwxy$,
the subtree for root A_j has yield w, and
the subtree for root A_i has yield vwx.

We can replace the subtree for root A_i with the subtree for root A_j, giving a tree with yield uwy (corresponding to the case $i = 0$), which also belongs to L.

We can replace the subtree for root A_j with the subtree for root A_i, giving a tree with yield uv^2wx^2y (corresponding to the case $i = 2$), which also belongs to L.

Etc.

Examples

While CFLs can match two sub-strings for (in) equality of length, they cannot match three such sub-strings.

Example 1: Consider $L = \{0^m 1^m 2^m \mid m \geq 1\}$.

Pick n of the pumping lemma. Pick $z = 0^n 1^n 2^n$. Break z into $uvwxy$, with $|vwx| \leq n$ and $vx \neq \epsilon$. Hence vwx cannot involve both 0s and 2s, since the last 0 and the first 2 are at least $n+1$ positions apart. There are two cases:

- vwx has no 2s. Then vx has only 0s and 1s. Then uwy, which would have to be in L, has n 2s, but fewer than n 0s or 1s.

- vwx has no 0s. Analogous.

Hence L is *not* a CFL.

Examples (continued)

CFLs cannot match two pairs of sub-strings of equal lengths if the pairs interleave.

Example 2: Consider $L = \{0^i 1^j 2^i 3^j \mid i, j \geq 1\}$.

Pick n of the pumping lemma. Pick $z = 0^n 1^n 2^n 3^n$. Break z into $uvwxy$, with $|vwx| \leq n$ and $vx \neq \epsilon$. Then vwx contains one or two different symbols. In both cases, the string uwy cannot be in L.

CFLs cannot match two sub-strings of arbitrary length over an alphabet of at least two symbols.

Example 3: Consider $L = \{ww \mid w \in \{0, 1\}^*\}$.

Pick n of the pumping lemma. Pick $z = 0^n 1^n 0^n 1^n$. In all cases, the string uwy cannot be in L.