Context Free Grammars - Lecture 9

James Marshall

More parse tree examples

<EXPR> → <EXPR>+<EXPR> | <EXPR>×<EXPR> | (<EXPR>) | a

Draw a parse tree for the string a+a×a

Definition - Ambiguous Grammar

A string is derived *ambiguously* in a CFG if there is more than one *leftmost derivation* of it. A CFG is *ambiguous* if it generates at least one string via an ambiguous derivation.

Definition - Inherently Ambiguous Language

If a language can *only* be generated by an ambiguous grammar, that language is said to be *inherently ambiguous*.

Example: $\{a^i b^j c^k \mid i = j \text{ or } j = k\}$

Definition - Chomsky Normal Form

A CFG is in *Chomsky Normal Form* if every rule has the form
(1) $A \rightarrow BC$, or
(2) $A \rightarrow a$
where (3) a is any terminal except ε, and A, B and C are any variables, with the important exception that (4) B and C may not be the start variable. If A is the start variable we may also have the rule $A \rightarrow \varepsilon$

Theorem

A CFG in Chomsky Normal Form exists for any CFL.

Proof Sketch (by construction)

Any grammar can be converted into Chomsky Normal Form by the following process:

1. Given existing start variable S add a new start variable S_0 and the rule $S_0 \rightarrow S$, to satisfy condition (4).
2. For any rule $A \rightarrow \varepsilon$ where A is not the start variable, to satisfy condition (3) For *every* occurence of A on the right-hand side of a rule, add a new rule with that occurence deleted. For rules of form $R \rightarrow A$ add rule $R \rightarrow \varepsilon$ *unless* that rule has already been deleted.
3. For any rules $A \rightarrow B$ and $B \rightarrow u$ (where u is a string of variables and terminals), delete $A \rightarrow B$ (to satisfy condition (1)) and add $A \rightarrow u$ (*unless* $A \rightarrow u$ was a rule already deleted in this stage).
4. Convert all remaining rules into the form of (1) or (2) by chaining rules together, *e.g.*

$A \rightarrow BCD$ becomes
$A \rightarrow BA_1$
$A_1 \rightarrow CD$ (satisfying (1))
and $D \rightarrow Ef$ becomes
$D \rightarrow EF$
$F \rightarrow f$ (satisfying (2))

Example

Write the following grammar in Chomsky Normal Form:
<EXPR> → <EXPR>+<EXPR> | (<EXPR>) | a

Theorem

If G is a CFG in Chomsky Normal Form, then for any string $w \in L(G)$ of length $n \geq 1$, there are $2n - 1$ steps in any derivation of w.

Proof Sketch