Ambiguity

Sometimes grammar can generate the same String in several different ways.

Such strings will have several different parse trees.

In some applications, this may be undesirable.

If a grammar generates a string in several different ways, we say that string is derived ambignously in that grammar.

If grammar generates some string ambignously, we say that the grammar is ambiguous.

Example:

$$\langle Expr \rangle \rightarrow \langle Expr \rangle + \langle Expr \rangle \mid \langle Expr \rangle \times \langle Expr \rangle \mid (\langle Expr \rangle) \mid a$$

$V = \{\langle Expr \rangle\}$ $\Sigma = \{a, (,), +, \times\}$

$S = \langle Expr \rangle$

a + a x a try to generate this

$\langle E \rangle \Rightarrow \langle E \rangle + \langle E \rangle \Rightarrow \langle E \rangle + \langle E \rangle \times \langle E \rangle$

$\Downarrow$

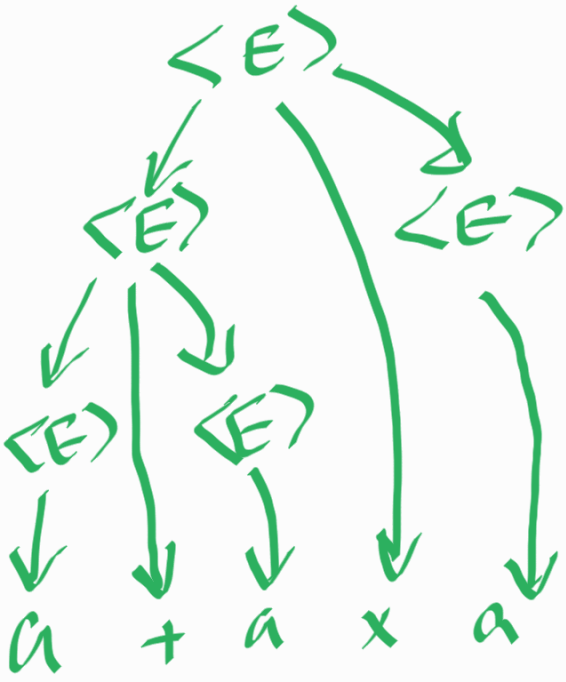
$a + a \times a$

$\langle E \rangle \Rightarrow \langle E \rangle \times \langle E \rangle \Rightarrow \langle E \rangle + \langle E \rangle \times \langle E \rangle$

$\Downarrow$

$a + a \times a$

⟨E⟩
⟨E⟩
⟨E⟩
⟨E⟩
⟨E⟩
⟨E⟩
a
+
a
+
a


⟨E⟩
⟨E⟩
⟨E⟩
⟨E⟩
⟨E⟩
a
+
a
x
a

Chomsky Normal Form (CNF)

When working with CFG, it is often convenient to have them in simplified form.

Very useful when we have algorithms working on CFGs.

Definition

Chomsky Normal Form (CNF)

Context free grammar is in chomsky normal form if every rule is in the form of:

- $A \longrightarrow BC$

$A \longrightarrow a$

a is a terminal, and A, B, C are any variable - except that B & C may not be start variable.

we also allow $S \longrightarrow \varepsilon$ (only for start variable) rules.

Theorem 2.9

Any context-free language is generated by a context-free grammar in Chomsky Normal Form.

Proof: (By construction)

Instead going through detailed proof, we will go through an example to understand this.

$S \to ASA \mid aB$

$A \to B \mid S$

$B \to b \mid \varepsilon$

Let's convert this to CNF.

1. Add a new starting variable.
   (This guarantees that start variable does not occur on the right hand side)

2. Take care of ε-rules.
   we remove $A \to \varepsilon$ rules where A is not the start variable.

for each occurence of A on the right hand side of a rule, we add a new rule with that occurrence deleted.

Ex $R \to uAvAw$ is replaced with

*$R \to uAvAw \mid uvAw \mid uAvw \mid uvw$

$R \to A \Rightarrow R \to \varepsilon$ (only if we did not remove this prev)

3. Handle all unit rules.

(i) we remove unit rules $A \to B$

(ii) when we see $B \to u$, we add $A \to u$ unless we have already removed it earlier.

u is a string of variables and terminals.

4. Convert all remaining rules into proper form.

(i) we replace each rule in the form of

$A \to u_1 u_2 \cdots u_k$, where $k \geq 3$, and u_i is a variable or terminal symbol, with rules

$A \to u_1 A_1 \quad A_1 \to u_2 A_2 \cdots A_{k-2} \to u_{k-1} u_k$

We replace any terminal u_i in the preceding rule(s) with the new variable U_i and $U_i \to u_i$

$A \longrightarrow u_1 u_2 u_3 \cdots u_k$

$A \longrightarrow u_1 A_1$

$A_1 \longrightarrow u_2 u_3 u_4 \cdots u_k$

$A_1 \longrightarrow u_2 A_2$

$A_2 \longrightarrow u_3 u_4 \cdots u_k$

Let's do an example

$S \to ASA \mid aB$

$A \to B \mid S$

$B \to b \mid \varepsilon$

Adding a new starting variable.

$S_0 \to S$

$S \to ASA \mid aB$

$A \to B \mid S$

$B \to b \mid \varepsilon$

removing $B \to \varepsilon$

$S_0 \to S$

$S \to ASA \mid aB \mid a$

$A \to B \mid S \mid \varepsilon$

$B \to b$

removing $A \to \varepsilon$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid SA \mid AS \mid S \mid aB \mid a$

$A \rightarrow B \mid S$

$B \rightarrow b$

removing $A \rightarrow B$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid SA \mid AS \mid S \mid aB \mid a$

$A \rightarrow S \mid b$

$B \rightarrow b$

remove $A \rightarrow S$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid SA \mid AS \mid S \mid aB \mid a$

$A \rightarrow b \mid ASA \mid SA \mid AS \mid aB \mid a$

$B \rightarrow b$

remove $S \to S$

$S_0 \to S$

$S \to ASA \mid SA \mid AS \mid aB \mid a$

$A \to b \mid ASA \mid SA \mid AS \mid aB \mid a$

$B \to b$

$S_0 \to ASA \mid SA \mid AS \mid aB \mid a$

$S \to ASA \mid SA \mid AS \mid aB \mid a$

$A \to b \mid ASA \mid SA \mid AS \mid aB \mid a$

$B \to b$

$S_0 \to AU_1 \mid SA \mid AS \mid aB \mid a$

$U_1 \to SA$

$S \to ASA \mid SA \mid AS \mid aB \mid a$

$A \to b \mid ASA \mid SA \mid AS \mid aB \mid a$

$B \to b$

$S_0 \rightarrow AU_1 \mid SA \mid AS \mid aB \mid a$

$U_1 \rightarrow SA$

$S \rightarrow AU_2 \mid SA \mid AS \mid aB \mid a$

$U_2 \rightarrow SA$

$A \rightarrow b \mid ASA \mid SA \mid AS \mid aB \mid a$

$B \rightarrow b$

$S_0 \rightarrow AU_1 \mid SA \mid AS \mid aB \mid a$

$U_1 \rightarrow SA$

$S \rightarrow AU_2 \mid SA \mid AS \mid aB \mid a$

$U_2 \rightarrow SA$

$A \rightarrow b \mid AU_3 \mid SA \mid AS \mid aB \mid a$

$U_3 \rightarrow SA$

$B \rightarrow b$

$S_0 \to AU_1 \mid SA \mid AS \mid U_4 B \mid a$

$U_4 \to a$

$U_1 \to SA$

$S \to AU_2 \mid SA \mid AS \mid U_5 B \mid a$

$U_5 \to a$

$U_2 \to SA$

$A \to b \mid AU_3 \mid SA \mid AS \mid U_6 B \mid a$

$U_6 \to a$

$U_3 \to SA$

$B \to b$

$S_0 \to AU_1 \mid SA \mid AS \mid U_4 B \mid a$

$U_4 \to a$

$U_1 \to SA$

$S \to AU_1 \mid SA \mid AS \mid U_4 B \mid a$

$A \to b \mid AU_1 \mid SA \mid AS \mid U_4 B \mid a$

$B \to b$

context free grammar

"context free" means that the application of a production rule depends only on the non-terminal being replaced, not on its surrounding

$\alpha A \beta \longrightarrow \alpha \gamma \beta$

A can only be replaced by γ only if it appears between α & β.