Standardized tests

Chomsky and Greibach normal forms



CS235 Languages and Automata

Department of Computer Science
Wellesley College

Normal forms

- A CFG is in Chomsky normal form (CNF) if all the productions are of the form

  $A \to BC$ or $A \to a$,

  where $A, B, \in N$ and $a \in \Sigma$.

- A CFG is in Greibach normal form (GNF) if all productions are of the form

  $A \to aB_1B_2 \ldots B_k$

  for some $k \geq 0$, where $A, B_1, \ldots, B_k \in N$ and $a \in \Sigma$.

For example,

- $S \to AB \mid AC \mid SS$
  $C \to SB$
  $A \to [$
  $B \to]$

- $S \to [B \mid [SB \mid [B\,S \mid [SB\,S$
  $B \to]$

*Compare with $G_{balanced} = S \to [S] \mid SS \mid \lambda$

WLOG we may assume …

Theorem 21.2. For any CFG G, there is a CFG G' in Chomsky normal form and a CFG G'' in Greibach normal form such that

$$L(G'') = L(G') = L(G) - \{\lambda\}.$$

Getting rid of □- and unit productions

- Derivations in grammars without these productions always make progress toward their sentential goals.
- At each step sentential forms either get longer and/or a nonterminal is replaced by a terminal.



When good rules go bad

Lemma 21.3. For any CFG $G = (N, \Sigma, P, S)$, there is a CFG G' with no □- or unit productions such that $L(G') = L(G) - \{\square\}$.

Proof. Let $\hat{P}$ be the smallest set of production containing P and closed under the two rules

(a) if $A \to \square B \square$ and $B \to \square$ are in $\hat{P}$, then $A \to \square\square$ is in $\hat{P}$; and

(b) if $A \to B$ and $B \to \square$ are in $\hat{P}$, then $A \to \square$ is in $\hat{P}$.

Let $\hat{G} = (N, \Sigma, \hat{P}, S)$ and note that since $P \subseteq \hat{P}$, $L(G) \subseteq L(\hat{G})$. In fact, $L(G) = L(\hat{G})$.

Chomsky normal form

We derive a CNF grammar for the set of nonnull strings of balanced brackets starting from the grammar

$S \to [S] \mid SS \mid \varepsilon$



Greibach normal form

- Start with a grammar

  $G = (N, \Sigma, P, S)$

  in Chomsky normal form.
- For $\alpha, \beta \in (N \cup \Sigma)^*$, write

  $\alpha \overset{L}{\Rightarrow}_G \beta$

  if β can be derived from α by a sequence of steps in which productions can be applied only to the leftmost symbol.
- For all $A \in N$ and $a \in \Sigma$,

  $R_{A,a} = \{b \in N^* \mid A \overset{L}{\Rightarrow}_G a\beta\}$



GNF: The saga continues



- Since $R_{A,a} = \{b\alpha \mid N^* \mid A \overset{L}{\underset{G}{\Rightarrow}} a\alpha\}$ is regular, it has a strongly right-linear grammar $G_{A,a}$. Let $T_{A,a}$ be the start symbol of $G_{A,a}$.
- Form the grammar G_1 by adding all the nonterminals of the grammar $G_{A,a}$ to G. Let S be the start symbol of G_1. Productions in G_1 are of the form:

  $X \to b, \quad X \to \lambda, \quad X \to BY,$

  where $b \in \Sigma$ and $B \in N$.
- Let G_2 be the grammar obtained from G_1 by replacing productions of the form $X \to BY$ by $X \to bT_{B,b}Y$ for all $b \in \Sigma$. Productions in G_2 have the form

  $X \to b, \quad X \to \lambda, \quad X \to bT_{B,b}Y,$

  where $b \in \Sigma$.
- Finally obtain G_3 by removing λ-productions from G_2.

GNF for nonnull balanced brackets

- $S \to AB \mid AC \mid SS, \quad C \to SB, \quad A \to [, \quad B \to]$

First compute regular sets $R_{D,d}$

- $S \to AB \mid AC \mid SS, \quad C \to SB, \quad A \to [, \quad B \to]$

- $R_{S,[} = (B + C)S^*$,
  $R_{C,[} = (B + C)S^*B$,
  $R_{A,[} = R_{b,[} = \{\varepsilon\}$,
  others are $\varnothing$.

Compute strongly right-linear grammars

- $S \to AB \mid AC \mid SS, \quad C \to SB, \quad A \to [, \quad B \to]$

- $R_{S,[} = (B + C)S^*$,
  $R_{C,[} = (B + C)S^*B$,
  $R_{A,[} = R_{b,]} = \{\varepsilon\}$,
  others are $\varnothing$.

- $T_{S,[} \to BX \mid CX, \quad X \to SX \mid \varepsilon$
  $T_{C,[} \to BY \mid CY, \quad Y \to SY \mid BZ, \quad Z \to \varepsilon$
  $T_{A,[} \to \varepsilon$
  $T_{B,]} \to \varepsilon$

G_2: Combine grammars & make substitutions

- $S \to AB \mid AC \mid SS, \quad C \to SB, \quad A \to [, \quad B \to]$

- $T_{S,[} \to BX \mid CX, \quad X \to SX \mid \lambda$
  $T_{C,[} \to BY \mid CY, \quad Y \to SY \mid BZ, \quad Z \to \lambda$
  $T_{A,[} \to \lambda$
  $T_{B,]} \to \lambda$

- $S \to [T_{A,[}B \mid [T_{A,[}C \mid [T_{S,[}S, \quad C \to [T_{S,[}B, \quad A \to [, B \to],$
  $T_{S,[} \to]T_{B,]}X \mid [T_{C,[}X, \quad X \to [T_{S,[}X \mid \lambda,$
  $T_{C,[} \to]T_{B,]}Y \mid [T_{C,[}Y, \quad Y \to [T_{S,[}Y \mid]T_{B,]}Z, \quad Z \to \lambda,$
  $T_{A,[} \to \lambda$
  $T_{B,]} \to \lambda$

G_3: Remove λ-transitions

- $S \to AB \mid AC \mid SS, \quad C \to SB, \quad A \to [, \quad B \to]$

- $S \to [T_{A,[}B \mid [T_{A,[}C \mid [T_{S,[}S, \quad C \to [T_{S,[}B, \quad A \to [, B \to],$
  $T_{S,[} \to]T_{B,]}X \mid [T_{C,[}X, \quad X \to [T_{S,[}X \mid \lambda,$
  $T_{C,[} \to]T_{B,]}Y \mid [T_{C,[}Y, \quad Y \to [T_{S,[}Y \mid]T_{B,]}Z, \quad Z \to \lambda,$
  $T_{A,[} \to \lambda$
  $T_{B,]} \to \lambda$

- $S \to [B \mid [C \mid [T_{S,[}S, \quad C \to [T_{S,[}B, \quad A \to [, B \to],$
  $T_{S,[} \to]X \mid [T_{C,[}X \mid [T_{C,[}, \quad X \to [T_{S,[}X \mid [T_{S,[},$
  $T_{C,[} \to]Y \mid [T_{C,[}Y, \quad Y \to [T_{S,[}Y \mid] \mid].$