Normal Forms

<u>Recall</u>: We showed that CFLs = languages accepted by PDAs
Used normal forms for CFGs to show this equivalence.

<u>Today</u>: How to convert any grammar into these normal forms.

We used Greibach Normal Form, where every rule is of the form

$$A \longrightarrow c B_1 \cdots B_k, \text{ where}$$

$k \geq 0$, $c \in T$, $A, B_1, \ldots, B_k \in NT$.

There is another, Chomsky Normal Form, where every rule is either of the form $A \to c$ or $A \longrightarrow BC$, where $c \in T$, $A, B, C \in NT$.

<u>Thm</u>: For every CFG G, there is a CFG G_1 in Chomsky Normal Form, and a CFG G_2 in Greibach Normal Form s.t.

$$\mathcal{L}(G_1) = \mathcal{L}(G_2) = \mathcal{L}(G) \setminus \{\varepsilon\}.$$

How do we show this?

We begin by getting rid of rules of the form

$A \to \varepsilon$ ε-production and

$A \to B$ unit-production.

We might need to replace these with other rules!

If I remove $S \to \varepsilon$ from $S \to \varepsilon \mid aSb \mid bSa \mid SS$ without adding any other rules in its place,

what language does my resultant grammar generate?

How can I fix this?

Claim: For any CFG $G = (NT, T, R, S)$, there is a CFG G' with no ε- or unit-productions such that $\mathcal{L}(G') = \mathcal{L}(G) \setminus \{\varepsilon\}$

Proof: Consider a set of rules $\hat{R}$ s.t. $R \subseteq \hat{R}$, and

① if $A \to \alpha B \beta \in \hat{R}$ and $B \to \varepsilon \in \hat{R}$, then $A \to \alpha\beta \in \hat{R}$ and

② if $A \to B \in \hat{R}$ and $B \to \gamma \in \hat{R}$, then $A \to \gamma \in \hat{R}$.

where $A, B \in NT$, $\alpha, \beta, \gamma \in (NT \cup T)^*$.

* How do we know that $\hat{R}$ does not keep growing infinitely large?

$\hat{R}$ is finite, and can be constructed inductively from R.

Now consider $\hat{G} = (NT, T, \hat{R}, S)$. Easy to show $\mathcal{L}(\hat{G}) = \mathcal{L}(G) \setminus \{\varepsilon\}$

Exercise!

We now have to show that if $x \in \Sigma^+$ is derived using some sequence of application of rules in $\hat{R}$, it does not use any ε- or unit-productions. If we show this, we can throw out those rules and get the desired G'.

Consider $x \neq \varepsilon$, and consider a shortest derivation for x in $\hat{G}$.

(i) Suppose $B \rightarrow \varepsilon$ is used at some point in this derivation.

$$S \xrightarrow{}^{*} \alpha B \gamma \longrightarrow \alpha\gamma \xrightarrow{}^{*} x$$

At least one of α and γ must be $\neq \varepsilon$.

So, there must be some $\delta, \eta, \zeta, \sigma$ s.t.

$$S \xrightarrow{i} \delta\eta B \zeta\sigma \xrightarrow{j} \alpha B \gamma \xrightarrow{1} \alpha\gamma \xrightarrow{k} x$$

But how do we get $\delta\eta B\zeta\sigma$? Via some rule of the form $A \to \eta B\zeta$.

$$\text{So, } S \xrightarrow{i-1} \delta A\sigma \xrightarrow{1} \delta\eta B\zeta\sigma \xrightarrow{j} \alpha B\gamma \xrightarrow{1} \alpha\gamma \xrightarrow{k} x. \qquad i+j+1+k$$

$$S \xrightarrow{0} S \xrightarrow{1} aSb \xrightarrow{0} aSb \xrightarrow{1} ab \xrightarrow{0} ab$$

But since we have $A \to \eta B\zeta$ and $B \to \varepsilon$ in $\hat{R}$, by ①,
$A \to \eta\zeta \in \hat{R}$.

So we can construct a shorter derivation for x, as follows:

$$S \xrightarrow{i-1} \delta A\sigma \xrightarrow{1} \delta\eta\zeta\sigma \xrightarrow{j} \alpha\gamma \xrightarrow{k} x. \qquad i+j+k$$

This cuts out at least one rule application from a minimal one for x.
Contradiction!

(ii) Suppose $A \to B$ is used at some point in a minimal derivation of x.

$$S \xrightarrow{}^{*} \alpha A \beta \longrightarrow \alpha B \beta \xrightarrow{}^{*} x$$

Since $x \in T^*$ and $B \in NT$, B must have been disposed of via some rule, of the form $B \to \gamma$. So, there must be γ, δ, η s.t.

$$S \xrightarrow{i} \alpha A \beta \xrightarrow{j} \delta A \eta \xrightarrow{1} \delta B \eta \xrightarrow{1} \delta \gamma \eta \xrightarrow{k} x \qquad i+1+j+1+k$$

But since $A \to B \in \hat{R}$ and $B \to \gamma \in \hat{R}$, $A \to \gamma \in \hat{R}$.

So, we get a shorter derivation for x as follows:

$$S \xrightarrow{i} \alpha A \beta \xrightarrow{j} \delta A \eta \xrightarrow{1} \delta \gamma \eta \xrightarrow{k} x. \qquad i+1+j+k$$

which contradicts the minimality of the above derivation of x.