# Context-Free Grammars (part II.5)

Prof. Dan A. Simovici

UMB

## Theorem

*For every context-free grammar $G$, there is a context-free, $\lambda$-free grammar $G'$ such that $L(G') = L(G) - \{\lambda\}$.*

## Proof.

Let $G = (A_N, A_T, S_0, P)$ be a context-free grammar. Consider the sequence $U_0, \ldots, U_m, \ldots$ of subsets of $A_N$ defined by

$$\begin{aligned} U_0 &= \{X \mid X \in A_N \text{ and } X \to \lambda \in P\}, \\ U_{m+1} &= U_m \cup \{X \in A_N \mid X \to \alpha \in P \text{ for some } \alpha \in U_m^*\}, \end{aligned}$$

for $m \in \mathbb{N}$.

Since $U_0 \subseteq U_1 \subseteq \cdots \subseteq A_N$, there is $k \in \mathbb{N}$ such that $U_k = U_{k+1}$. □

# (Proof cont'd)

A simple argument (by induction on $h \geq 1$) shows that $U_k = U_{k+h}$ for every $h \geq 1$.

The base step is immediate.

Suppose that $U_k = U_{k+h}$ and let $X \in U_{k+h+1}$. If $X \in U_{k+h}$, then $X \in U_k$ by the inductive hypothesis. Otherwise, there is a production $X \to \alpha \in P$ such that $\alpha \in U^*_{k+h}$. By the inductive hypothesis, $\alpha \in U^*_k$, so $X \in U_{k+1} = U_k$. Therefore, $U_{k+h+1} = U_k$.

We claim that $X \overset{+}{\underset{G}{\Rightarrow}} \lambda$ if and only if $X \in U_k$.

We prove by strong induction on $p \geq 1$ that if $X \overset{p}{\underset{G}{\Rightarrow}} \lambda$, then $X \in U_k$.
For $p = 1$, if $X \underset{G}{\Rightarrow} \lambda$, then $X \in U_0$ and $U_0 \subseteq U_k$.

# (Proof cont'd)

Suppose that the statement is true for derivations $X \overset{+}{\underset{G}{\Rightarrow}} \lambda$ of length no greater than $p$ and let $X \overset{p+1}{\underset{G}{\Rightarrow}} \lambda$. The first production applied in this derivation must have the form $X \rightarrow X_{i_1} \cdots X_{i_q}$; therefore, we have

$$X_{i_1} \cdots X_{i_q} \overset{p}{\underset{G}{\Rightarrow}} \lambda.$$

Hence, $X_{i_\ell} \overset{p_\ell}{\underset{G}{\Rightarrow}} \lambda$, where $p_\ell \leq p$ for $1 \leq \ell \leq q$. By the inductive hypothesis, we have $X_{i_\ell} \in U_k$, so $X_{i_1} \cdots X_{i_q} \in (U_k)^*$, which implies $X \in U_{k+1} = U_k$.

# (Proof cont'd)

Conversely, it is easy to prove (by induction on $n$) that for every $X \in U_n$ we have $X \overset{+}{\underset{G}{\Rightarrow}} \lambda$. We leave this argument to the reader. From this it follows that if $\theta \in U_k^*$, then $\theta \overset{*}{\underset{G}{\Rightarrow}} \lambda$.

Consider now the set of productions $P'$, where

$$\begin{aligned} P' \;=\; & \{X \to \alpha' \mid \alpha' \neq \lambda, \text{ there is } X \to \alpha \in P \text{ and } \alpha' \text{ is obtained} \\ & \text{from } \alpha \text{ by erasing 0 or more symbols from } U_k\}. \end{aligned}$$

# (Proof cont'd)

If $G'$ is the context-free grammar $G' = (A_N, A_T, S_0, P')$, then $L(G') = L(G) - \{\lambda\}$. Indeed, suppose that $X \overset{p}{\underset{G'}{\Rightarrow}} \gamma$. Clearly, $\gamma \neq \lambda$ since $G'$ has no erasure productions. We prove, by strong induction on $p$, that we have $X \overset{*}{\underset{G}{\Rightarrow}} \gamma$.

For $p = 0$, the statement is trivially true.

# (Proof cont'd)

Assume that it holds for derivations of length less than or equal to $p$, and let $X \underset{G'}{\overset{p+1}{\Rightarrow}} \gamma$. If the first production applied in this derivation is $X \to X_{i_0} \cdots X_{i_{h-1}}$, then $\gamma = \gamma_0 \ldots \gamma_{h-1}$, where $X_{i_j} \underset{G'}{\overset{p_j}{\Rightarrow}} \gamma_j$, $p_j \leq p$, for $0 \leq j \leq h-1$. By the inductive hypothesis we have $X_{i_j} \underset{G}{\overset{*}{\Rightarrow}} \gamma_j$ for $0 \leq j \leq h-1$.

# (Proof cont'd)

Furthermore, assume that the production $X \rightarrow X_{j_0} \cdots X_{j_{h-1}}$ was obtained from the production $X \rightarrow \theta_0 X_{j_0} \theta_1 \cdots X_{j_{h-1}} \theta_h$ from $P$, where $\theta_0, \ldots, \theta_h \in (U_k)^*$. Our previous discussion allows us to infer the existence of the derivations $\theta_q \overset{*}{\underset{G}{\Rightarrow}} \lambda$ for $0 \leq q \leq h$. By combining the derivations obtained above, we have

$$\begin{aligned} X & \underset{G}{\Rightarrow} \theta_0 X_{j_0} \theta_1 \cdots X_{i_{h-1}} \theta_h \\ & \overset{*}{\underset{G}{\Rightarrow}} X_{j_0} \cdots X_{j_{h-1}} \\ & \overset{*}{\underset{G}{\Rightarrow}} \gamma_0 \cdots \gamma_{h-1} = \gamma. \end{aligned}$$

This implies $L(G') \subseteq L(G) - \{\lambda\}$.

# (Proof cont'd)

To prove the converse inclusion, consider a derivation $X \overset{p}{\underset{G}{\Rightarrow}} \gamma$, where $\gamma \neq \lambda$.
We claim that $X \overset{*}{\underset{G'}{\Rightarrow}} \gamma$. The argument is by strong induction on $p \geq 0$.
The case $p = 0$ is trivially true. Assume that the statement holds for derivations of length of no more than $p$, and let $X \overset{p+1}{\underset{G}{\Rightarrow}} \gamma$, where $\gamma \neq \lambda$.
Let $\beta = X_{j_0} \cdots X_{j_{l-1}}$ be the word that follows $X$ in the previous derivation, that is, $X \underset{G}{\Rightarrow} \beta \overset{p}{\underset{G}{\Rightarrow}} \gamma$.

# (Proof cont'd)

We can write:

$$\gamma = \gamma_0 \cdots \gamma_{l-1},$$

where $X_{j_m} \overset{p_m}{\underset{G}{\Rightarrow}} \gamma_m$ and $p_m \leq p$ for $0 \leq m \leq l-1$.

If $\gamma_m \neq \lambda$, by the inductive hypothesis, we have $X_{j_m} \overset{*}{\underset{G'}{\Rightarrow}} \gamma_m$. On the other hand, if $\gamma_m = \lambda$, we have $X_{j_m} \in U_k$. Let

$$\{h_0, \ldots, h_{q-1}\} = \{h \mid 0 \leq h \leq l-1 \text{ and } \gamma_h \neq \lambda\}.$$

The definition of $P'$ implies that we have the production $X \to X_{j_{h_0}} \cdots X_{j_{h_{q-1}}}$ in $P'$. Therefore,

$$X \underset{G'}{\Rightarrow} X_{j_{h_0}} \cdots X_{j_{h_{q-1}}} \overset{*}{\underset{G'}{\Rightarrow}} \gamma_{h_0} \cdots \gamma_{h_{q-1}} = \gamma.$$

This implies $L(G) - \{\lambda\} \subseteq L(G')$.

### Theorem

*If $G$ is a context-free grammar, then there is an equivalent context-free grammar $G'$ such that one of the following two cases occurs:*

1. *if $\lambda \notin L(G)$, then $G'$ is $\lambda$-free;*
2. *if $\lambda \in L(G)$, then $G'$ contains a unique erasure production $S' \to \lambda$, where $S'$ is the start symbol of $G'$, and $S'$ does not occur in any right member of any production of $G'$.*

# Proof

We have shown that for every context-free grammar $G$ there is a context-free, $\lambda$-free grammar $G_1$ such that $L(G_1) = L(G) - \{\lambda\}$. If $\lambda \notin L(G)$, then the grammars $G$ and $G_1$ are equivalent, and we can define $G'$ as $G_1$. This proves the first case of this theorem.
If $\lambda \in L(G)$, by the same theorem, we have the context-free, $\lambda$-free grammar $G_1 = (A_N, A_T, S_1, P)$ such that $L(G_1) = L(G) - \{\lambda\}$. Define the grammar $G'$ by

$$G' = (A_N \cup \{S'\}, A_T, S', \{S' \to S_1, S' \to \lambda\} \cup P),$$

where $S'$ is a new nonterminal symbol (i.e., that $S' \notin A_N$). It is immediate that $G'$ satisfies the conditions of the second case of this Theorem and that $L(G') = L(G_1) \cup \{\lambda\} = L(G)$.

## Example

Let $G = (\{S, X, Y, Z\}, \{a, b\}, S, \{S \to XYZ, X \to YZ, X \to aYb, X \to a, Y \to \lambda, Y \to b, Z \to \lambda, Z \to c\})$ be a context-free grammar that contains erasure productions. The sequence of subsets of $\{S, X, Y, Z\}$ is

$$U_0 = \{Y, Z\}, U_1 = \{Y, Z, X\}, U_2 = \{Y, Z, X, S\}, U_3 = U_2.$$

Therefore, the set of productions $P'$ is given by

$$\begin{aligned} P' \ = \ & \{S \to XYZ, S \to YZ, S \to XZ, S \to XY, S \to X, S \to Y, \\ & S \to Z, X \to YZ, X \to Y, X \to Z, X \to aYb, X \to ab, \\ & X \to a, Y \to b, Z \to c\} \end{aligned}$$

Observe that the productions of $P'$ are obtained by erasing zero, one, or more of the symbols $X, Y, Z$ from the rules of $P$.

The previous theorem shows that it is possible to limit the erasure productions in context-free grammars that generate a language $L$ to a single production that has the start symbol as its left member, without restricting the generality.

**Corollary**

*Every context-free language is a context-sensitive language; in other words,* $\mathcal{L}_2 \subseteq \mathcal{L}_1$*.*

**Proof.**

This is an immediate consequence of a previous theorem and the definitions of $\mathcal{L}_1$ and $\mathcal{L}_2$. □

**Definition**

Let $G = (A_N, A_T, S, P)$ be a context-free grammar. A *chain production* is a production $X \to Y$, where $X, Y \in A_N$.

**Theorem**

*Let $G = (A_N, A_T, S, P)$ be a context-free grammar. There is a context-free grammar $G_1$ such that $G_1$ is equivalent to $G$ and $G_1$ does not contain chain productions.*

# Proof

We assume initially that $G$ is $\lambda$-free. Let $X$ be a nonterminal symbol. To eliminate productions of the form $X \to Y$ consider the following sequence of sets:

$$\begin{aligned} U_0^X &= \{X\} \\ U_{n+1}^X &= U_n^X \cup \{Z \in A_N \mid Y \to Z \in P \text{ for some } Y \in U_n^X\} \end{aligned}$$

It is clear that the sequence $U_0^X, \ldots, U_n^X, \ldots$ is an increasing sequence of subsets of $A_N$. The finiteness of $A_N$ implies the existence of a number $i$ such that $U_i^X = U_{i+1}^X$. Then, by induction on $\ell \geq 1$, we can easily prove that $U_i^X = U_{i+\ell}^X$ for $\ell \geq 1$.

We shall prove that $U_i^X = \{Z \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} Z\}$.

A straightforward argument by induction on $n$ shows that

$U_n^X \subseteq \{Z \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} Z\}$ for $n \in \mathbb{N}$. In particular,

$U_i^X \subseteq \{Z \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} Z\}$.

# (Proof cont'd)

To prove the converse inclusion, we prove that if a derivation $X \overset{k}{\underset{G}{\Rightarrow}} Z$, then $Z \in U_i^X$. The argument is by induction on $k$. For $k = 0$, $Z = X$, and $Z \in U_0^X \subseteq U_i^X$, so the conclusion follows. Suppose that the statement holds for derivations of length $k$, and let $X \overset{k+1}{\underset{G}{\Rightarrow}} Z'$. Since the grammar has no erasure rules, we can write $X \overset{k}{\underset{G}{\Rightarrow}} Z \underset{G}{\Rightarrow} Z'$. By the inductive hypothesis, $Z \in U_i^X$; the existence of the production $Z \rightarrow Z'$ implies that $Z \in U_{i+1}^X = U_i^X$. Thus, $\{Z \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} Z\} \subseteq U_i^X$.

# (Proof cont'd)

Denote the set $\{Z \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} Z\}$ by $U_*^X$. The context-free grammar $G_1 = (A_N, A_T, S, P_1)$ is defined by

$$P_1 = \{X \to \alpha \mid Z \to \alpha \in P \text{ for some } Z \in U_*^X \text{ and } \alpha \notin A_N\}.$$

It is clear that the grammar $G_1$ has no chain productions and is equivalent to $G$.

If $G$ is not $\lambda$-free, then there exists an equivalent context-free grammar $G' = (A_N \cup \{S'\}, A_T, S', P' \cup \{S' \to \lambda\})$ where $S' \to \lambda$ is the unique erasure production of $G'$, and $S'$ does not occur in any right member of any production of $G'$. The grammar $G'' = (A_N \cup \{S'\}, A_T, S', P')$ generates the language $L(G) - \{\lambda\}$. By applying the previous construction to $G''$ we obtain the grammar $G''_1 = (A_N \cup \{S'\}, A_T, S', P''_1)$ that has no chain rules and for which $L(G''_1) = L(G) - \{\lambda\}$. Then, the desired grammar $G_1$ is given by

$$G_1 = (A_N \cup \{S'\}, A_T, S', P''_1 \cup \{S' \to \lambda\}),$$

where $S_1$ is a new start symbol.

**Example**

The grammar

$$\begin{aligned} G \;=\; & (\{S, X, Y\}, \{a, b, c\}, S, \{S \to X, S \to aX, X \to Y, \\ & X \to bY, S \to a, X \to b, Y \to c\}) \end{aligned}$$

is $\lambda$-free and contains some chain productions.

# (Example cont'd)

We have $U_0^S = \{S\}$, $U_1^S = \{S, X\}$, $U_2^S = \{S, X, Y\}$, and $U_2^S = U_3^S = \cdots$, so $U_*^S = \{S, X, Y\}$. Similar computations give $U_*^X = \{X, Y\}$ and $U_*^Y = \{Y\}$. The grammar

$$\begin{aligned} G_1 &= (\{S, X, Y\}, \{a, b, c\}, S, \{S \to aX, S \to bY, S \to a, S \to b, \\ &\quad S \to c, X \to c, X \to bY, X \to b, Y \to c\}). \end{aligned}$$

is equivalent to $G$ and has no chain productions.

Let $G = (A_N, A_T, S, P)$ be a context-free grammar, and let $X$ be a nonterminal symbol. Denote by $L(G, X)$ the set of terminal words that can be generated from $X$ in the grammar $G$,, that is,

$$L(G, X) = \{x \in A_T^* \mid X \overset{*}{\underset{G}{\Rightarrow}} x\}.$$

Clearly, we have $L(G, S) = L(G)$.

### Definition

Let $G = (A_N, A_T, S, P)$ be a context-free grammar. A symbol $s \in A_N \cup A_T$ is *accessible* if it occurs in a word $\alpha \in (A_N \cup A_T)^*$ such that $S \overset{*}{\underset{G}{\Rightarrow}} \alpha$.

A symbol $X \in A_N$ is *productive* if $L(G, X) \neq \emptyset$.

**Theorem**

*Let $G = (A_N, A_T, S, P)$ be a context-free grammar. There is a construction of an equivalent grammar $G' = (A'_N, A_T, S, P')$ such that $P' = \emptyset$ if $L(G) = \emptyset$, and if $L(G) \neq \emptyset$, then every symbol in $A'_N$ is productive.*

# Proof

Define the sequence $U_0, \ldots, U_n, \ldots$ of subsets of $A_N$ by

$$\begin{aligned} U_0 &= \{X \in A_N \mid X \to u \in P \text{ for some } u \in A_T^*\} \\ U_{n+1} &= U_n \cup \{X \in A_N \mid X \to \alpha \in P \text{ for some } \alpha \in (U_n \cup A_T)^*\} \end{aligned}$$

Note that $U_0 \subseteq U_1 \subseteq \cdots \subseteq U_n \subseteq \cdots \subseteq A_N$. Therefore, there is $i$ such that $U_i = U_{i+1}$. An easy argument by induction on $k$ shows that $U_i = U_{i+k}$ for $k \geq 1$. This part of the proof is left to the reader.
We claim that

$$\{X \in A_N \mid L(G, X) \neq \emptyset\} = U_i.$$

# (Proof cont'd)

If $n = 0$, the conclusion follows from the definition of $U_0$. Suppose that the inclusion holds for $U_n$ and let $Y \in U_{n+1}$. If $Y \in U_n$ the conclusion is immediate. Otherwise, there is a production $Y \to \alpha$, where $\alpha = w_0 Z_0 w_1 Z_1 \cdots w_{p-1} Z_{p-1} w_p$, where $w_i \in A_T^*$ for $0 \leq i \leq p$ and $Z_j \in U_n$ for $0 \leq j \leq p-1$. By the inductive hypothesis, we have the derivations $Z_j \overset{*}{\underset{G}{\Rightarrow}} z_j$, where $z_j \in A_T^*$ for $0 \leq j \leq p-1$. Thus, we obtain the derivation

$$Y \underset{G}{\Rightarrow} w_0 Z_0 w_1 Z_1 \cdots w_{p-1} Z_{p-1} w_p \overset{*}{\underset{G}{\Rightarrow}} w_0 z_0 w_1 z_1 \cdots w_{p-1} z_{p-1} w_p \in A_T^*,$$

which gives the desired conclusion. In particular,

$$U_i \subseteq \{X \in A_N \mid X \overset{*}{\underset{G}{\Rightarrow}} u \text{ for some } u \in A_T^*\}.$$

# (Proof cont'd)

To prove the converse inclusion we prove by strong induction on $m \geq 1$ that $X \overset{m}{\underset{G}{\Rightarrow}} u$ for $u \in A_T^*$ implies $X \in U_{m-1}$. The basis case, $m = 1$, is immediate.

Suppose that the statement holds for derivations of length less than or equal to $m$ and consider a derivation $X \overset{m+1}{\underset{G}{\Rightarrow}} u$ for $u \in A_T^*$. If we write the first step of this derivation, we obtain

$$X \underset{G}{\Rightarrow} w_0 Z_0 w_1 Z_1 \cdots w_{p-1} Z_{p-1} w_p \overset{m}{\underset{G}{\Rightarrow}} u,$$

where $w_0, \ldots, w_{p-1}, u \in A_T^*$, and $Z_0, \ldots, Z_{p-1} \in A_N$.

# (Proof cont'd)

The word $u$ can be written as $u = w_0 z_0 w_1 z_1 \cdots w_{p-1} z_{p-1} w_p$, where $Z_j \overset{\ell_j}{\underset{G}{\Rightarrow}} z_j$, $\ell_j \leq m$ for $0 \leq j \leq p-1$. By the inductive hypothesis, we have $Z_j \in U_{\ell_j - 1} \subseteq U_{m-1}$, so $w_0 Z_0 w_1 Z_1 \cdots w_{p-1} Z_{p-1} w_p \in (U_{m-1} \cup A_T)^*$. Thus, $X \in U_m$.

# (Proof cont'd)

Since $U_m \subseteq U_i$ for every $m \in \mathbb{N}$ and $m \geq 1$, we obtain the converse inclusion and, therefore, the desired equality.
Note that $S \in U_i$ if and only if $L(G) \neq \emptyset$. Define the set of productions $P'$ by

$$P' = \begin{cases} \emptyset \text{ if } S \notin U_i \\ \{X \to \alpha \mid \alpha \in (U_i \cup A_T)^* \text{ and } X \to \alpha \in P\} \text{ otherwise.} \end{cases}$$

# (Proof cont'd)

Since $P' \subseteq P$ it follows that $L(G') \subseteq L(G)$. Conversely, if $u \in L(G)$, then $S \overset{*}{\underset{G}{\Rightarrow}} u$. Let $X \rightarrow \alpha$ be a production that occurs in this derivation. We have

$$S \overset{*}{\underset{G}{\Rightarrow}} \beta X \gamma \overset{*}{\underset{G}{\Rightarrow}} \beta \alpha \gamma \overset{*}{\underset{G}{\Rightarrow}} u.$$

Therefore, every nonterminal symbol that occurs in $\alpha$ must be productive. This allows us to conclude that $\alpha \in (U_i \cup A_T)^*$, hence $X \rightarrow \alpha \in P'$. Since every production used in the derivation $S \overset{*}{\underset{G}{\Rightarrow}} u$ belongs to $P'$, it follows that $u \in L(G')$, so $L(G) \subseteq L(G')$.

### Corollary

*The emptiness of the language* $L(G)$ *generated by a context-free grammar* $G = (A_N, A_T, S, P)$ *is decidable.*

### Proof.

Note that the start symbol $S$ of a context-free grammar $G$ is productive if and only if $L(G) \neq \emptyset$. Therefore, in order to decide if $L(G) = \emptyset$, it suffices to compute the set $U_i$. Then, $L(G) = \emptyset$ if and only if $S \notin U_i$. □

Example

Let $G = (\{S, X, Y, Z\}, \{a, b\}, S, \{S \to YZ, S \to XY, S \to XZ, Z \to ab, Y \to bc\})$ be a context-free grammar. The sequence $U_0, U_1, \ldots$ is given by $U_0 = \{Y, Z\}$, $U_1 = \{S, Y, Z\}$, $U_1 = U_2 = \cdots$. Therefore, the grammar $G' = (\{S, Y, Z\}, \{a, b\}, S, \{S \to YZ, Z \to ab, Y \to bc\})$ has only productive symbols and is equivalent to $G$.

**Theorem**

*Let $G = (A_N, A_T, S, P)$ be a context-free grammar. There exists an equivalent context-free grammar $G' = (A'_N, A_T, S, P')$ such that every production of $P'$ that contains a terminal symbol is of the form $X \to a$.*

# Proof

Consider the alphabet $A' = \{X_a \mid a \in A_T\}$ that contains a symbol $X_a$ for every terminal symbol $a$, where $A_N \cap A' = \emptyset$, and define $A'_N$ as $A'_N = A_N \cup A'$.
The productions of $P'$ are obtained by replacing each terminal symbol $a$ by the corresponding nonterminal $X_a$ and by adding the productions $X_a \to a$ for $a \in A_T$. The set of productions $P'$ satisfies the requirements of the theorem, and the resulting grammar is clearly of the same type as $G$.

# Proof (cont'd)

Let $u = a_{i_0} \cdots a_{i_{n-1}} \in L(G)$. The definition of the grammar $G'$ implies that $S \overset{*}{\underset{G'}{\Rightarrow}} X_{a_{i_0}} \cdots X_{a_{i_{n-1}}}$. By using the productions $X_a \rightarrow a$ we obtain

$$S \overset{*}{\underset{G'}{\Rightarrow}} a_{i_0} \cdots a_{i_{n-1}},$$

so $a_{i_0} \cdots a_{i_{n-1}} \in L(G')$. Thus, $L(G) \subseteq L(G')$.

# Proof (cont'd)

To prove the converse inclusion, $L(G') \subseteq L(G)$, consider a morphism $h : (A'_N \cup A_T)^* \longrightarrow (A_N \cup A_T)^*$ defined by $h(X_a) = a$ for $a \in A_T$ and $h(Y) = Y$ for every $Y \in A_N \cup A_T$. We claim that if $\alpha \underset{G'}{\Rightarrow} \beta$ for some $\alpha, \beta \in (A'_N \cup A_T)^*$, then $h(\alpha) \overset{*}{\underset{G}{\Rightarrow}} h(\beta)$. Indeed, if a production of the form $X \rightarrow a$ was used in $\alpha \underset{G'}{\Rightarrow} \beta$, then $h(\alpha) = h(\beta)$.

# Proof (cont'd)

If another kind of production was used, then $h(\alpha) \underset{G}{\Rightarrow} h(\beta)$, so in any case, $h(\alpha) \overset{*}{\underset{G'}{\Rightarrow}} h(\beta)$. Let now $v \in L(G')$. We have $S \overset{*}{\underset{G'}{\Rightarrow}} v$, so $S = h(S) \overset{*}{\underset{G}{\Rightarrow}} h(v) = v$, which implies $v \in L(G)$. Therefore, $L(G) = L(G')$.

**Example**

Recall the context-free grammar $G = (A_N, A_T, S_0, P)$, where $A_N = \{S_0, S_1, S_2\}$, $A_T = \{a, b\}$, and $P$ contains the following productions:

$$S_0 \to aS_2, S_0 \to bS_1, S_1 \to a, S_1 \to aS_0,$$
$$S_1 \to bS_1S_1, S_2 \to b, S_2 \to bS_0, S_2 \to aS_2S_2.$$

# (Example cont'd)

To limit productions that contain terminal symbols to productions of the form $X_a \to a$ add the non-terminal symbols $X_a$ and $X_b$. The set of productions becomes

$$S_0 \to X_a S_2, S_0 \to X_b S_1, S_1 \to X_a, S_1 \to X_a S_0,$$
$$S_1 \to X_b S_1 S_1, S_2 \to X_b, S_2 \to X_b S_0, S_2 \to X_a S_2 S_2,$$
$$X_a \to a, X_b \to b.$$