# Grammars
# (part II)

Prof. Dan A. Simovici

UMB

1. Length-Increasing vs. Context-Sensitive Grammars
2. Closure Properties of Classes of Languages
3. Properties of Type-2 Grammars

**Theorem**

*The class $\mathcal{L}_1$ equals the class of length-increasing languages.*

# Proof

Clearly, every type-1 grammar is length-increasing. Therefore, as we observed earlier, $\mathcal{L}_1$ is included in the class of length-increasing languages. To prove the converse inclusion, consider a length-increasing grammar $G = (A_N, A_T, S, P)$. We can assume that every production of $P$ that contains a terminal symbol is of the form $X \to a$.

# Proof (cont'd)

Let $\pi : \alpha \to \beta \in P$ be a production such that $\alpha = X_0 \cdots X_{n-1}$, $\beta = Y_0 \cdots Y_{m-1}$. If $n = 1$, then $\pi$ is already a context-sensitive production. Therefore, suppose that $2 \leq n \leq m$. By hypothesis, $X_0, \ldots, X_{n-1}, Y_0, \ldots, Y_{m-1}$ are nonterminals. Consider $n$ new nonterminals $Z_0^\pi, \ldots, Z_{n-1}^\pi$ and the set of productions $P_\pi$ that consists of:

$$
\begin{aligned}
& X_0 \cdots X_{n-1} \to Z_0^\pi X_1 \cdots X_{n-1} \\
& Z_0^\pi X_1 \cdots X_{n-1} \to Z_0^\pi Z_1^\pi X_2 \cdots X_{n-1} \\
& \vdots \\
& Z_0^\pi Z_1^\pi \cdots Z_{n-2}^\pi X_{n-1} \to Z_0^\pi Z_1^\pi \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \\
& Z_0^\pi Z_1^\pi \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \to Y_0 Z_1^\pi \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \\
& Y_0 Z_1^\pi \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \to Y_0 Y_1 \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \\
& \vdots \\
& Y_0 Y_1 \cdots Z_{n-1}^\pi Y_n \cdots Y_{m-1} \to Y_0 Y_1 \cdots Y_{n-1} Y_n \cdots Y_{m-1}.
\end{aligned}
$$

The set $P_\pi$ consists of context-sensitive productions.

# Proof (cont'd)

Let $P_1$ be the set of productions that consists of $\bigcup\{P_{\alpha\to\beta} \mid 2 \leq |\alpha| \leq |\beta|\}$ and the productions of the form $X \to \beta$ or $S \to \lambda$ (whenever such productions belong to $P$). Consider the context-sensitive grammar $G_1 = (A_N \cup A, A_T, S, P_1)$, where $A$ consists of all new nonterminal symbols introduced when the production sets $P_\pi$ were constructed.

# Proof (cont'd)

Note that a derivation step in $G$ that consists of the application of a production $\pi : \alpha \to \beta$ with $2 \leq |\alpha| \leq |\beta|$ corresponds to the successive application of the productions of $P_\pi$ in $G_1$; conversely, the productions of $P_\pi$ can be applied only in this order in a derivation in $G_1$, and they simulate a step in a derivation in $G$ that makes use of the production $\pi$. A step that uses a production $X \to \beta$ or $S \to \lambda$ is the same in both $G$ and $G_1$. Thus, $S \overset{*}{\underset{G}{\Rightarrow}} x$ if and only if $S \overset{*}{\underset{G_1}{\Rightarrow}} x$, so $L(G) = L(G_1)$. This shows that the class of length-increasing languages is included in $\mathcal{L}_1$.

### Example

Let

$$G' = (\{S, X, Y, X_a, X_b, X_c\}, \{a, b, c\}, S, P')$$

be the length-increasing previously constructed grammar. The set of productions $P_1$ consists of:

- The context-sensitive productions in $P'$, namely:

$$\begin{array}{lclclcl} \pi'_0 & : & S \to X_a X_b X_c, & \pi'_1 & : & S \to X_a X X_b X_c, \\ \pi'_7 & : & X_a \to a, & \pi'_8 & : & X_b \to b \\ \pi'_9 & : & X_c \to c & & & \end{array}$$

# Example (cont'd)

Each remaining production in $P'$ generates the following productions:

- For $\pi'_2 : XX_b \to X_bX$, include the productions:

$$\begin{array}{lcl} \pi'_{10} & : & XX_b \to Z_0X_b \\ \pi'_{11} & : & Z_0X_b \to Z_0Z_1 \\ \pi'_{12} & : & Z_0Z_1 \to X_bZ_1 \\ \pi'_{13} & : & X_bZ_1 \to X_bX \end{array}$$

- For $\pi'_3 : XX_c \to YX_bX_cX_c$, include the productions:

$$\begin{array}{lcl} \pi'_{14} & : & XX_c \to Z_2X_c \\ \pi'_{15} & : & Z_2X_c \to Z_2Z_3X_cX_c \\ \pi'_{16} & : & Z_2Z_3X_cX_c \to YZ_3X_cX_c \\ \pi'_{17} & : & YZ_3X_cX_c \to YX_bX_cX_c \end{array}$$

# Example (cont'd)

- For $\pi'_4 : X_b Y \to Y X_b$, include the productions:

$$\begin{array}{lcl} \pi'_{18} & : & X_b Y \to Z_4 Y \\ \pi'_{19} & : & Z_4 Y \to Z_4 Z_5 \\ \pi'_{20} & : & Z_4 Z_5 \to Y Z_5 \\ \pi'_{21} & : & Y Z_5 \to Y X_b \end{array}$$

- For $\pi'_5 : X_a Y \to X_a X_a X$, include the productions:

$$\begin{array}{lcl} \pi'_{22} & : & X_a Y \to Z_6 Y \\ \pi'_{23} & : & Z_6 Y \to Z_6 Z_7 X \\ \pi'_{24} & : & Z_6 Z_7 X \to X_a Z_7 X \\ \pi'_{25} & : & X_a Z_7 X \to X_a X_a X \end{array}$$

- Finally, for $\pi'_6 : X_a Y \to X_a X_a$, include the productions:

$$\begin{array}{lcl} \pi'_{26} & : & X_a Y \to Z_8 Y \\ \pi'_{27} & : & Z_8 Y \to Z_8 Z_9 \\ \pi'_{28} & : & Z_8 Z_9 \to X_a Z_9 \\ \pi'_{29} & : & X_a Z_9 \to X_a X_a \end{array}$$

This gives the context-sensitive grammar:

$$\begin{array}{rcl} G_1 & = & (\{S, X, Y, X_a, X_b, X_c\} \cup \{Z_i \mid 0 \leq i \leq 10\}, \{a, b, c\}, S, \\ & & \{\pi'_0, \pi'_1, \pi'_7, \pi'_8, \pi'_9\} \cup \{\pi'_j \mid 10 \leq j \leq 29\}. \end{array}$$

that generates the language $L = \{a^n b^n c^n \mid n \in \mathbb{N}\}$. This implies that $L$ is a context-sensitive language.

### Theorem

*The classes $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2$ are closed with respect to the reversal operation; in other words, if $L \in \mathcal{L}_i$, then $L^R \in \mathcal{L}_i$ for $i \in \{0, 1, 2\}$.*

# Proof

Let $G = (A_N, A_T, S, P)$ be a grammar, and let $P^R = \{\alpha^R \to \beta^R \mid \alpha \to \beta \in P\}$.
Define $G^R = (A_N, A_T, S, P^R)$. If $G$ is of type $i$, then so is $G^R$ for $i \in \{0, 1, 2\}$.
Further, we have $\gamma \overset{*}{\underset{G}{\Rightarrow}} \gamma'$ if and only if $\gamma^R \overset{*}{\underset{G^R}{\Rightarrow}} \gamma'^R$, as can be shown by induction on the length of the two derivations. Thus, $S \overset{*}{\underset{G}{\Rightarrow}} w$ if and only if $S = S^R \overset{*}{\underset{G^R}{\Rightarrow}} w^R$. This shows that $L(G^R) = L(G)^R$, so $L(G)^R$ has the same type as the language $L(G)$.

**Theorem**

*Each of the classes $\mathcal{L}_i$ of Chomsky's hierarchy contains the class of finite languages, for $i \in \{0, 1, 2\}$.*

**Proof.**

Let $L = \{u_0, \ldots, u_{n-1}\}$ be a finite, nonempty language over an alphabet $A$. The grammar $G = (\{S\}, A, S, \{S \to u_0, \ldots, S \to u_{n-1}\})$ is of type 3 and, therefore, of type 2, 1, and 0. If $L = \emptyset$, then $L$ is generated by the grammar $G = (\{S\}, A, S, \{S \to S\})$ that is, again, of type 3. □

### Theorem

*If $L$ is a language of type $i$, where $i \in \{0, 1\}$, then so is the language $K$, where $K = L - \{\lambda\}$.*

### Proof.

If $L$ is a type-1 language, then there exists a type-1 grammar $G = (A_N, A_T, S, P)$ such that $L(G) = L$. If $\lambda \in L$, the production $S \rightarrow \lambda$ belongs to $P$, and $S$ does not occur in the right member of any production. The language $K$ is generated by the type-1 grammar $G' = (A_N, A_T, S, P - \{S \rightarrow \lambda\})$. If $\lambda \notin L$, we have $K = L$.
If $L$ is of type 0 and $\lambda \in L$, let $P'$ be the set of productions obtained from $P$ by replacing all erasure productions $\alpha \rightarrow \lambda$ with $Y\alpha \rightarrow Y$ and $\alpha Y \rightarrow Y$ for every $Y \in A_N \cup A_T$. The grammar $(A_N, A_T, S, P')$ generates the language $K$, so $K$ is a type-0 language. □

### Theorem

*$\mathcal{L}_i$ is closed with respect to union, for $i \in \{0, 1, 2, 3\}$.*

### Proof.

Suppose that $L, L'$ are two languages of type $i$ that are generated by the grammars $G = (A_N, A_T, S, P)$ and $G' = (A'_N, A_T, S', P')$, respectively, where $A_N \cap A'_N = \emptyset$.
Consider the grammar
$G_\cup = (A_N \cup A'_N \cup \{S_0\}, A_T, S_0, P \cup P' \cup \{S_0 \to S, S_0 \to S'\})$, where $S_0$ is a new nonterminal symbol such that $S_0 \notin A_N \cup A'_N$. Note that the grammar $G_\cup$ is of the same type $i$ as the grammars $G$ and $G'$. To complete the proof, we need to show that $L \cup L' = L(G_\cup)$. □

# Proof (cont'd)

Consider the grammar $G_\cup = (A_N \cup A'_N \cup \{S_0\}, A_T, S, P \cup P' \cup \{S_0 \to S, S_0 \to S'\})$, where $S_0$ is a new nonterminal symbol such that $S_0 \notin A_N \cup A'_N$. Note that the grammar $G_\cup$ is of the same type $i$ as the grammars $G$ and $G'$. To complete the proof, we need to show that $L \cup L' = L(G_\cup)$.

Let $x \in L \cup L'$. If $x \in L$, then $S \overset{*}{\underset{G}{\Rightarrow}} x$, so $S_0 \underset{G_\cup}{\Rightarrow} S \overset{*}{\underset{G_\cup}{\Rightarrow}} x$ which shows that $x \in L(G_\cup)$. The case when $x \in L'$ is entirely similar and is left to the reader. Thus, $L \cup L' \subseteq L(G_\cup)$.

# Proof (cont'd)

Conversely, suppose that $x \in L(G_\cup)$. We have $S_0 \overset{*}{\underset{G_\cup}{\Rightarrow}} x$. If the first production applied in this derivation is $S_0 \to S$, then the derivation can be written as $S_0 \underset{G_\cup}{\Rightarrow} S \overset{*}{\underset{G_\cup}{\Rightarrow}} x$. The last part of this derivation $S \overset{*}{\underset{G_\cup}{\Rightarrow}} x$ uses only productions from $P$ since $A_N \cap A'_N = \emptyset$ implies $P \cap P' = \emptyset$. Therefore, we have $S \overset{*}{\underset{G}{\Rightarrow}} x$, so $x \in L(G)$. Similarly, if the first production applied is $S_0 \to S'$, then $x \in L(G')$. Therefore, $L(G_\cup) \subseteq L(G) \cup L(G')$, hence $L(G_\cup) = L(G) \cup L(G')$.

In Slides 20–30 we show that the classes $\mathcal{L}_i$ are closed with respect to the $*$ operation for $0 \leqslant i \leqslant 3$.

## Lemma

*The classes $\mathcal{L}_0$ and $\mathcal{L}_1$ are closed with respect to the $*$ operation.*

## Proof.

We must prove that if $L \in \mathcal{L}_i$, then $L^* \in \mathcal{L}_i$ for $i \in \{0, 1\}$. Let us assume initially that $\lambda \notin L$.

Let $G = (A_N, A_T, S, P)$ be a grammar of type 0 or type 1 that generates the language $L$. We can assume that terminal symbols do not occur in the left member of any production of $P$.

Let $S_0, S_1$ be new symbols such that $S_0, S_1 \notin A_N$, and let $G_* = (A_N \cup \{S_0, S_1\}, A_T, S_0, P_*)$ be the grammar whose set of productions is

$$\begin{aligned} P_* &= P \cup \{S_0 \to \lambda, S_0 \to S, S_0 \to S_1 S\} \cup \\ &\quad \{S_1 a \to S_1 S a \mid a \in A_T\} \cup \{S_1 a \to S a \mid a \in A_T\}. \end{aligned}$$

□

# Proof (cont'd)

If $G$ is of type 1, then so is $G_*$.
We have $L^n \subseteq L(G_*)$ for $n \in \mathbb{N}$.
Let $x \in L^n$. If $n = 0$ we have $x = \lambda$, and $x \in L(G_*)$ because $S_0 \to \lambda$ is in $P_*$. Suppose now that $x = x_0 x_1 \ldots x_{n-1}$, where $x_0, \ldots, x_{n-1} \in L$ for $n \geq 1$. We have $S \overset{*}{\underset{G}{\Rightarrow}} x_j$ for $0 \leq j \leq n-1$, and the same derivations are valid in $G_*$ because $P \subset P_*$. Thus, we can put together the following derivation in the grammar $G_*$:

$$
\begin{aligned}
S_0 & \underset{G_*}{\Rightarrow} S_1 S \overset{*}{\underset{G_*}{\Rightarrow}} S_1 x_{n-1} \\
& \underset{G_*}{\Rightarrow} S_1 S x_{n-1} \overset{*}{\underset{G_*}{\Rightarrow}} S_1 x_{n-2} x_{n-1} \\
& \vdots \\
& \underset{G_*}{\Rightarrow} S_1 S x_2 \cdots x_{n-1} \overset{*}{\underset{G_*}{\Rightarrow}} S_1 x_1 \cdots x_{n-1} \\
& \underset{G_*}{\Rightarrow} S x_1 \cdots x_{n-1} \overset{*}{\underset{G_*}{\Rightarrow}} x_0 x_1 \cdots x_{n-1} = x.
\end{aligned}
$$

# Proof (cont'd)

This proves that $x \in L(G_*)$. Thus, $L^* = \bigcup_{n\in\mathbb{N}} L^n \subseteq L(G_*)$. Observe that the "regeneration" of the symbol $S$ is made possible in the above derivation by the fact that the words $x_j$ are not null (which puts $S_1$ adjacent with terminal symbols, thereby allowing the application of the productions $S_1 a \rightarrow S_1 S a$).

# Proof (cont'd)

Conversely, let $x \in L(G_*)$. There exists a derivation $S_0 \overset{*}{\underset{G_*}{\Rightarrow}} x$. If we write this derivation explicitly,

$$S_0 \underset{G_*}{\Rightarrow} \alpha_1 \underset{G_*}{\Rightarrow} \cdots \underset{G_*}{\Rightarrow} \alpha_{n-1} = x,$$

we have to consider three cases:

`Case 1:` If the first production applied in this derivation is $S_0 \to \lambda$, then $x = \lambda$, so $x \in L^*$.

`Case 2:` If the first production applied is $S_0 \to S$, then we have $S \overset{*}{\underset{G_*}{\Rightarrow}} x$, and the same derivation is valid in $G$, which shows that $x \in L(G) = L$.

# Proof (cont'd)

`Case 3:` If the first production applied is $S_0 \to S_1 S$, then each word $\alpha_\ell$ in the above derivation falls into one of the following two cases:

`Case 3a:` $\alpha_\ell$ is of the form $S_1 \beta_1 \cdots \beta_k$, where $k \geq 1$, $S \overset{*}{\underset{G}{\Rightarrow}} \beta_q$ for $1 \leq q \leq k$, and the first symbol of each of the words $\beta_2, \ldots, \beta_k$ is a terminal.

`Case 3b:` $\alpha_\ell$ is of the form $\beta_0 \beta_1 \cdots \beta_k$, where $k \geq 1$, the first symbol of each $\beta_p$ is a terminal for $p \geq 1$, and $S \overset{*}{\underset{G}{\Rightarrow}} \beta_p$ for $0 \leq p \leq k$.

# Proof (cont'd)

This argument is by induction on $\ell \geq 1$. It is clear that $\alpha_1$ falls in Case 3a. Suppose that $\alpha_h$ satisfies the conditions of either Case 3a or Case 3b. Since the productions of $P$ do not have terminal symbols in their left members, their application may involve only one word $\beta_i$. This guarantees that if $\alpha_{h+1}$ was obtained through the application of a production in $P$, then $\alpha_{h+1}$ satisfies one of the conditions of the previous cases. The same conclusion can be reached if $\alpha_{h+1}$ was obtained by applying a production of the form $S_1 a \longrightarrow S_1 S a$ or $S_1 a \longrightarrow S a$.

# Proof (cont'd)

The existence of the derivation $S_0 \overset{*}{\underset{G_*}{\Rightarrow}} x \in A_T^*$ implies that $x$ is a word of the form 3b. Therefore $x$ can be written as a product, $x = \beta_0 \cdots \beta_k$, where $k \geq 1$, and $S \overset{*}{\underset{G}{\Rightarrow}} \beta_p$ for $0 \leq p \leq k$. Since every word $\beta_0, \beta_1, \ldots, \beta_k \in A_T^*$ it follows that $\beta_0, \beta_1, \ldots, \beta_k \in L(G)$. Thus, $x \in (L(G))^{k+1} \subseteq L(G)^*$ which implies $L(G_*) \subseteq L(G)^*$.
If $\lambda \in L$, then consider the language $K = L - \{\lambda\}$. The language $K$ is of the same type as $L$ and, by the above argument, so is $K^*$. Since $K^* = L^*$ we obtain the desired closure property.

## Lemma

*The class $\mathcal{L}_2$ is closed with respect to the $*$ operation.*

## Proof.

Let $L$ be a context-free language generated by the type-2 grammar $G = (A_N, A_T, S, P)$. Suppose that $S_0$ is a new nonterminal symbol and consider the type-2 grammar $G_* = (A_N \cup \{S_0\}, A_T, S_0, P \cup \{S_0 \to \lambda, S_0 \to S_0 S\})$. It is easy to verify that $L(G_*) = L^*$, so $L^* \in \mathcal{L}_2$. □

**Lemma**

*The class $\mathcal{L}_3$ is closed with respect to the $*$ operation.*

**Proof.**

Let $L \in \mathcal{L}_3$ such that $L = L(G)$, where $G = (A_N, A_T, S, P)$ is a type-3 grammar. Define the set of productions $P_1 = \{X \to uS \mid X \to u \in P\}$. Consider the type-3 grammar

$$G_* = (A_N \cup \{S_0\}, A_T, S_0, P \cup P_1 \cup \{S_0 \to \lambda, S_0 \to S\}),$$

where $S_0$ be a new nonterminal symbol, $S_0 \notin A_N$. It is easy to verify that $L(G_*) = L^*$, so $L^* \in \mathcal{L}_3$. □

**Theorem**

*The classes $\mathcal{L}_i$ are closed with respect to the $*$ operation for $i \in \{0, 1, 2, 3\}$.*

**Proof.**

This follows from the previous lemmas. □

### Corollary

*The classes $\mathcal{L}_i$ are closed with respect to the $+$ operation for $i \in \{0, 1, 2, 3\}$.*

### Proof.

Let $\mathcal{L}_i$ be one of the classes of Chomsky's hierarchy, and let $L \in \mathcal{L}_i$. Note that

$$L^+ = \begin{cases} L^* & \text{if } \lambda \in L \\ L^* - \{\lambda\} & \text{if } \lambda \notin L. \end{cases}$$

In all cases $L^+ \in \mathcal{L}_i$. □

In the next few slides (Slides 32-37) we prepare the necessary results for proving that the classes $\mathcal{L}_i$ are closed with respect to the product operation for $0 \leqslant i \leqslant 3$.

## Lemma

*The classes $\mathcal{L}_0$ and $\mathcal{L}_2$ are closed with respect to the product operation.*

## Proof.

Let $L, L'$ be two languages of type $i$, and let $G = (A_N, A_T, S, P)$, $G' = (A'_N, A_T, S', P')$ be two grammars of type $i$ such that $L(G) = L$ and $L(G') = L'$, where $i \in \{0, 2\}$. Without any loss of generality, we can assume that $A_N \cap A'_N = \emptyset$.
If $S_0$ is a new symbol, $S_0 \notin A_N \cup A'_N$, then the grammar $G_p = (A_N \cup A'_N \cup \{S_0\}, A_T, S_0, P \cup P' \cup \{S_0 \to SS'\})$ is also of type $i$. We claim that $L(G_p) = LL'$. □

# Proof (cont'd)

Let $x \in LL'$. We can write $x = uv$ for some $u \in L$ and $v \in L'$. By hypothesis, $S \underset{G}{\overset{*}{\Rightarrow}} u$ and $S' \underset{G'}{\overset{*}{\Rightarrow}} v$, so

$$S_0 \underset{G_p}{\Rightarrow} SS' \underset{G_p}{\overset{*}{\Rightarrow}} uS' \underset{G_p}{\overset{*}{\Rightarrow}} uv = x.$$

Thus, $x \in L(G_p)$.

Conversely, suppose that $x \in L_p$. There is a derivation

$$S_0 \underset{G_p}{\Rightarrow} SS' \overset{*}{\underset{G_p}{\Rightarrow}} x.$$

Since $A_N$ and $A'_N$ are disjoint sets, the sets of productions $P$ and $P'$ are disjoint. Therefore, the productions of $G_p$ used to transform $S$ into a word over $A_T$ belong to $P$, while the ones used to rewrite $S'$ belong to $P'$. Thus, we can write $x = uv$, where $S \overset{*}{\underset{G}{\Rightarrow}} u$ and $S' \overset{*}{\underset{G'}{\Rightarrow}} v$, which implies $x \in LL'$.

## Lemma

*The class $\mathcal{L}_1$ is closed with respect to the product operation.*

## Proof.

Let $L, L'$ be two languages in $\mathcal{L}_1$. If neither $L$ nor $L'$ contains the null word, we may assume that both languages are generated by type-1 grammars that have no erasure rules. It is easy to see that, in this case, the construction of the grammar $G_p$ given in the proof of previous Lemma yields a type-1 grammar, so $LL'$ belongs to $\mathcal{L}_1$. □

# Proof cont'd

Suppose now that $\lambda \in L$ or $\lambda \in L'$. By The languages $L_1 = L - \{\lambda\}$ and $L_1' = L' - \{\lambda\}$ also belong to $\mathcal{L}_1$ and, by the previous argument $L_1 L_1' \in \mathcal{L}_1$. We need to consider the cases summarized below.

| Case | $\lambda \in L$ | $\lambda \in L'$ | $L$ | $L'$ | $LL'$ |
|---|---|---|---|---|---|
| 1 | yes | no | $L_1 \cup \{\lambda\}$ | $L_1'$ | $L_1 L_1' \cup L_1'$ |
| 2 | no | yes | $L_1$ | $L_1' \cup \{\lambda\}$ | $L_1 L_1' \cup L_1$ |
| 3 | yes | yes | $L_1 \cup \{\lambda\}$ | $L_1' \cup \{\lambda\}$ | $L_1 L_1' \cup L_1 \cup L_1' \cup \{\lambda\}$ |

In each case, we have $LL' \in \mathcal{L}_1$.

### Lemma

*The class $\mathcal{L}_3$ is closed with respect to the product operation.*

### Proof.

Let $L, L'$ be two languages in $\mathcal{L}_3$ and assume that $L, L'$ are generated by the grammars $G = (A_N, A_T, S, P)$ and $G' = (A'_N, A'_T, S', P')$, respectively. Without loss of generality, we may assume that $A_N \cap A'_N = \emptyset$; this also implies $P \cap P' = \emptyset$. Consider the set of productions $P_1 = \{X \to uY \mid X \to uY \in P\} \cup \{X \to uS' \mid X \to u \in P\}$, which is obtained from $P$ by replacing every production $X \to u$ by a production $X \to uS'$. □

# Proof cont'd

The type-3 grammar $G_1 = (A_N \cup A'_N, A_T \cup A'_T, S, P_1 \cup P')$ generates the language $LL'$. Indeed, if $x \in L(G)$ and $y \in L(G')$, then $S \overset{*}{\underset{G}{\Rightarrow}} x$ and $S' \overset{*}{\underset{G'}{\Rightarrow}} y$. Since $P$ was replaced by $P_1$ in $G_1$, we have $S \overset{*}{\underset{G_1}{\Rightarrow}} xS'$. Note that we also have $S' \overset{*}{\underset{G_1}{\Rightarrow}} y$. Combining the last two derivations we can write $S \overset{*}{\underset{G_1}{\Rightarrow}} xy$, so $LL' \subseteq L(G_1)$.

# Proof cont'd

Conversely, let $z \in L(G_1)$. We have $S \underset{G_1}{\overset{*}{\Rightarrow}} z$. This derivation begins with a symbol from $A_N$ and must eventually use a production from $P'$ since, otherwise, nonterminal symbols cannot be erased. Therefore, the last derivation can be written as

$$S \underset{G_1}{\overset{*}{\Rightarrow}} uS' \underset{G_1}{\overset{*}{\Rightarrow}} uv = z,$$

where $u \in A_T^*$ and $v \in {A'_T}^*$. This implies the existence of the derivations $S \underset{G_1}{\overset{*}{\Rightarrow}} uS'$ and $S' \underset{G_1}{\overset{*}{\Rightarrow}} v$. Note that the first derivation corresponds to $S \underset{G}{\overset{*}{\Rightarrow}} u$; the second corresponds to $S' \underset{G'}{\overset{*}{\Rightarrow}} v$. Thus, $u \in L(G)$, $v \in L(G')$, and this implies $z = uv \in LL'$. Therefore, $L(G_1) \subseteq LL'$.

**Theorem**

*Each of the classes $\mathcal{L}_i$ is closed with respect to the product operation.*

**Proof.**

Follows from previous lemmas. □

## Theorem

*Let $G = (A_N, A_T, S, P)$ be a context-free grammar. If*

$$X_0 \cdots X_{k-1} \overset{n}{\underset{G}{\Rightarrow}} \alpha,$$

*where $X_0, \ldots, X_{k-1} \in A_N \cup A_T$ and $\alpha \in (A_N \cup A_T)^*$, then we can write $\alpha = \alpha_0 \cdots \alpha_{k-1}$, where $X_i \overset{n_i}{\underset{G}{\Rightarrow}} \alpha_i$ for $0 \leq i \leq k-1$ and $\sum_{0 \leq i \leq k-1} n_i = n$.*

## Proof.

We use an argument by induction on $n$, $n \geq 0$. For $n = 0$, we have $\alpha_i = X_i$ for $0 \leq i \leq k-1$, and the statement is obviously true; in this case, $n_0 = \cdots = n_{k-1} = 0$. □

# Proof cont'd

Assume that the statement is true for derivations of length $n$, and let

$$X_0 \cdots X_{k-1} \overset{n+1}{\underset{G}{\Rightarrow}} \alpha.$$

If $X_0 \cdots X_{k-1} \overset{n}{\underset{G}{\Rightarrow}} \gamma \underset{G}{\Rightarrow} \alpha$, by the inductive hypothesis, we have $\gamma = \gamma_0 \cdots \gamma_{k-1}$, where $X_i \overset{n_i}{\underset{G}{\Rightarrow}} \gamma_i$ for $0 \leq i \leq k-1$ and $\sum\{n_i \mid 0 \leq i \leq k-1\} = n$.

# Proof cont'd

Let $Y \to \beta$ be the production applied in the last step $\gamma \underset{G}{\Rightarrow} \alpha$. $Y$ occurs in one of the words $\gamma_0, \ldots, \gamma_{k-1}$, say, $\gamma_j$. In this case, we can write $\gamma_j = \gamma_j' Y \gamma_j''$ and $\alpha$ can be written as $\alpha = \alpha_0 \cdots \alpha_{k-1}$, where $\alpha_i = \gamma_i$ for $0 \le i \le j-1$, and $j+1 \le i \le k-1$, $X_j \overset{n_j}{\underset{G}{\Rightarrow}} \gamma_j \underset{G}{\Rightarrow} \gamma_j' \beta \gamma_j'' = \alpha_j$, which proves the statement.

**Definition**

A derivation $\gamma_0 \underset{G}{\Rightarrow} \gamma_1 \underset{G}{\Rightarrow} \cdots \underset{G}{\Rightarrow} \gamma_n$ in a context-free grammar $G = (A_N, A_T, S, P)$ is *complete* if $\gamma_n \in A_T^*$.

Note that if $X_0 \cdots X_{k-1} \underset{G}{\Rightarrow} \cdots \underset{G}{\Rightarrow} \alpha$ is a complete derivation in $G$, then every derivation that results from "splitting" this derivation is also complete.

### Example

Let $G = (A_N, A_T, S_0, P)$ be a context-free grammar, 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

We prove that $L(G)$ consists of all nonnull words over $\{a, b\}$ that contain an equal number of $a$'s and $b$'s. Recall that $n_X(\alpha)$ is the number of occurrences of symbol $X$ in the word $\alpha$.
We will show by strong induction on $p$, $p \geq 1$, that

1. if $n_a(u) = n_b(u) = p$, then $S_0 \overset{*}{\underset{G}{\Rightarrow}} u$;
2. if $n_a(u) = n_b(u) + 1 = p$, then $S_1 \overset{*}{\underset{G}{\Rightarrow}} u$;
3. if $n_b(u) = n_a(u) + 1 = p$, then $S_2 \overset{*}{\underset{G}{\Rightarrow}} u$.

# Example cont'd

In the first case, for $p = 1$, we have either $u = ab$ or $u = ba$; hence, we have either $S_0 \underset{G}{\Rightarrow} aS_2 \underset{G}{\Rightarrow} ab$ or $S_0 \underset{G}{\Rightarrow} bS_1 \underset{G}{\Rightarrow} ba$.
For the second case, $u = a$, and we have $S_1 \underset{G}{\Rightarrow} a$; the third case, for $u = b$, is similar.

# Example cont'd

Suppose that the statement holds for $p \leq n$. Again, we consider three cases for the word $u$:

1. $n_a(u) = n_b(u) = n + 1$;
2. if $n_a(u) = n_b(u) + 1 = n + 1$;
3. if $n_b(u) = n_a(u) + 1 = n + 1$.

In the first case, we may have four situations:

$1_1$. $u = abt$, where $t \in \{a, b\}^*$ and $n_a(t) = n_b(t) = n$,

$1_2$. $u = bat$, where $t \in \{a, b\}^*$ and $n_a(t) = n_b(t) = n$,

$1_3$. $u = aav$ with $n_b(v) = n + 1$ and $n_a(v) = n - 1$, or

$1_4$. $u = bbw$ with $n_a(w) = n + 1$ and $n_b(w) = n - 1$.

# Example cont'd

By the inductive hypothesis, we have $S_0 \overset{*}{\underset{G}{\Rightarrow}} t$, and therefore, we obtain one of the following derivations:

$$S_0 \underset{G}{\Rightarrow} aS_2 \underset{G}{\Rightarrow} abS_0 \overset{*}{\underset{G}{\Rightarrow}} abt = u,$$
$$S_0 \underset{G}{\Rightarrow} bS_1 \underset{G}{\Rightarrow} baS_0 \overset{*}{\underset{G}{\Rightarrow}} bat = u,$$

for the cases $(1_1)$ and $(1_2)$, respectively.

# Example cont'd

On the other hand, if $u = aav$, we can write $v = v'v''$, where $v'$ is the shortest prefix of $v$, where the number of $b$s exceeds the number of $a$s. Clearly, we must have $n_b(v') = n_a(v') + 1 = n'$, and therefore, $n_b(v'') = n_a(v'') + 1 = n''$, where $n' + n'' = n + 1$. By the inductive hypothesis, we have $S_2 \overset{*}{\underset{G}{\Rightarrow}} v'$, $S_2 \overset{*}{\underset{G}{\Rightarrow}} v''$; hence,

$$S_0 \underset{G}{\Rightarrow} aS_2 \underset{G}{\Rightarrow} aaS_2S_2 \overset{*}{\underset{G}{\Rightarrow}} aav'v'' = u,$$

which concludes the argument for $(1_3)$. We leave to the reader the similar arguments for the remaining cases. This allows us to conclude that every word that contains an equal number of $a$'s and $b$'s belongs to $L(G)$.

# Example cont'd

To prove the reverse inclusion, we justify the following implications:

1. If $S_0 \underset{G}{\overset{n}{\Rightarrow}} \alpha$, then $n_a(\alpha) + n_{S_1}(\alpha) = n_b(\alpha) + n_{S_2}(\alpha)$.
2. If $S_1 \underset{G}{\overset{n}{\Rightarrow}} \alpha$, then $n_a(\alpha) + n_{S_1}(\alpha) = n_b(\alpha) + n_{S_2}(\alpha) + 1$.
3. If $S_2 \underset{G}{\overset{n}{\Rightarrow}} \alpha$, then $n_a(\alpha) + n_{S_1}(\alpha) + 1 = n_b(\alpha) + n_{S_2}(\alpha)$.

The proof is by strong induction on $n$, where $n \geq 1$. For $n = 1$, the verification is immediate. For instance, if $S_1 \underset{G}{\Rightarrow} \alpha$, we have $\alpha = a$, $\alpha = aS_0$, or $\alpha = bS_1S_1$; in every case, the equality is satisfied.

# Example cont'd

Suppose that the implications hold for derivations no longer than $n$.
If $S_0 \overset{n+1}{\underset{G}{\Rightarrow}} \alpha$, the first production applied in the derivation is $S_0 \to aS_2$ or $S_0 \to bS_1$. In the first case, we have $\alpha = a\beta$, where $S_2 \overset{n}{\underset{G}{\Rightarrow}} \beta$, and by the inductive hypothesis, we have $n_a(\beta) + n_{S_1}(\beta) + 1 = n_b(\beta) + n_{S_2}(\beta)$, so

$$\begin{aligned} n_a(\alpha) + n_{S_1}(\alpha) &= n_a(\beta) + 1 + n_{S_1}(\beta) \\ &= n_b(\beta) + n_{S_2}(\beta) \\ &= n_b(\alpha) + n_{S_2}(\alpha). \end{aligned}$$

The second case has a similar treatment.

# Example cont'd

If $S_1 \overset{n+1}{\underset{G}{\Rightarrow}} \alpha$, we have three possibilities.

(a) If the first production of the derivation is $S_1 \to a$, then $\alpha = a$ and the equality corresponding to this case is obviously satisfied.

(b) If the first production is $S_1 \to aS_0$, we can write $\alpha = a\beta$, where $S_0 \overset{n}{\underset{G}{\Rightarrow}} \beta$; hence, $n_a(\beta) + n_{S_1}(\beta) = n_b(\beta) + n_{S_2}(\beta)$, so

$$\begin{aligned} n_a(\alpha) + n_{S_1}(\alpha) &= n_a(\beta) + 1 + n_{S_1}(\beta) \\ &= n_b(\beta) + n_{S_2}(\beta) + 1 \\ &= n_b(\alpha) + n_{S_2}(\alpha) + 1. \end{aligned}$$

# Example cont'd

(c) If the derivation begins with $S_1 \to bS_1S_1$ we can write $\alpha = b\beta\gamma$, where $S_1 \overset{p}{\underset{G}{\Rightarrow}} \beta$ and $S_1 \overset{q}{\underset{G}{\Rightarrow}} \gamma$, where $p, q \leq n$. By the inductive hypothesis, $n_a(\beta) + n_{S_1}(\beta) = n_b(\beta) + n_{S_2}(\beta) + 1$, and $n_a(\gamma) + n_{S_1}(\gamma) = n_b(\gamma) + n_{S_2}(\gamma) + 1$. Consequently,

$$\begin{aligned} n_a(\alpha) + n_{S_1}(\alpha) &= n_a(\beta) + n_a(\gamma) + n_{S_1}(\beta) + n_{S_1}(\gamma) \\ &= n_b(\beta) + n_{S_2}(\beta) + 1 + n_b(\gamma) + n_{S_2}(\gamma) + 1 \\ &= n_b(\alpha) + n_{S_2}(\alpha) + 1. \end{aligned}$$

The case of the derivation $S_2 \overset{*}{\underset{G}{\Rightarrow}} \alpha$ can be treated in a similar manner.

# Example cont'd

Let $u \in L(G)$. From the existence of the derivation $S_0 \overset{*}{\underset{G}{\Rightarrow}} u$ we obtain $n_a(u) = n_b(u)$, which shows that $L(G) \subseteq \{x \in \{a, b\}^* \mid n_a(x) = n_b(x)\}$.