Context-Free languages
(part I)

Prof. Dan A. Simovici

UMB

1. Properties of Type-2 Grammars
2. Closure Properties of the class $\mathcal{L}_2$
3. Leftmost Derivations and Ambiguity

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, γ_j. In this case, we can write $\gamma_j = \gamma_j' Y \gamma_j''$ and α can be written as $\alpha = \alpha_0 \cdots \alpha_{k-1}$, where $\alpha_i = \gamma_i$ for $0 \leq i \leq j-1$, and $j+1 \leq i \leq k-1$, $X_j \overset{n_j}{\underset{G}{\Rightarrow}} \gamma_j \underset{G}{\Rightarrow} \gamma_j' \beta \gamma_j'' = \alpha_j$, 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 α.

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 $S_0 \stackrel{*}{\underset{G}{\Rightarrow}} t$, and therefore, we obtain one of the following derivations:

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

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 bs exceeds the number of as. 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 $S_0 \overset{n}{\underset{G}{\Rightarrow}} \alpha$, then $n_a(\alpha) + n_{S_1}(\alpha) = n_b(\alpha) + n_{S_2}(\alpha)$.
2. If $S_1 \overset{n}{\underset{G}{\Rightarrow}} \alpha$, then $n_a(\alpha) + n_{S_1}(\alpha) = n_b(\alpha) + n_{S_2}(\alpha) + 1$.
3. If $S_2 \overset{n}{\underset{G}{\Rightarrow}} \alpha$, 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 $L(G) \subseteq \{x \in \{a, b\}^* \mid n_a(x) = n_b(x)\}$.

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

$\mathcal{L}_2$ is closed with respect to union.

Proof.

Suppose that L, L' are two languages of type 2 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
$G_\cup = (A_N \cup A'_N \cup \{S_0\}, A_T, S, P \cup P' \cup \{S_0 \to S, S_0 \to S'\})$, 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 type 2 as the grammars G and G'. To complete the proof, we need to show that $L \cup L' = L(G_\cup)$. □

Proof (cont'd)

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 $L(G_\cup) = L(G) \cup L(G')$.

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 $L^* \in \mathcal{L}_3$. □

Lemma

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

Proof.

Let L, L' be two languages of type 2, and let $G = (A_N, A_T, S, P)$, $G' = (A'_N, A_T, S', P')$ be two grammars of type 2 such that $L(G) = L$ and $L(G') = L'$. Without 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, $S \overset{*}{\underset{G}{\Rightarrow}} u$ and $S' \overset{*}{\underset{G'}{\Rightarrow}} v$, so

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

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 $x \in LL'$.

Leftmost Derivations

Definition

Let $G = (A_N, A_T, S, P)$ be a context-free grammar.
A *leftmost derivation* is a derivation $\gamma_0 \Rightarrow \cdots \Rightarrow \gamma_n$ in G such that, if the production applied in deriving γ_{k+1} from γ_k is $X_k \rightarrow \beta_k$, then $\gamma_k = \gamma'_k X_k \gamma''_k$, $\gamma_{k+1} = \gamma'_k \beta_k \gamma''_k$ and $\gamma'_k \in A_T^*$.

- The words γ_k (for $0 \leq k \leq n$) are referred to as *left sentential forms*.
- If $\gamma_k = \gamma'_k X_k \gamma''_k$, where $\gamma'_k \in A_T^*$, then γ'_k is the *closed part* of γ_k, while $X_k \gamma''_k$ is the *open part* of γ_k.
- In a context-free grammar G,

$$\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$$

is a leftmost derivation if, at every step of this derivation, we always rewrite the leftmost nonterminal symbol.

Notations

- The existence of a leftmost derivation of length n in the context-free grammar G, $\gamma_0 \Rightarrow \gamma_1 \Rightarrow \ldots \Rightarrow \gamma_n$, will be denoted by $\gamma_0 \overset{n}{\underset{G,\text{left}}{\Rightarrow}} \gamma_n$.
- The existence of a leftmost derivation of any length of γ' from γ in the same grammar will be denoted by $\gamma \overset{*}{\underset{G,\text{left}}{\Rightarrow}} \gamma'$.
- The existence of a leftmost derivation of positive length of γ' from γ will be denoted by $\gamma \overset{+}{\underset{G,\text{left}}{\Rightarrow}} \gamma'$.

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)

The derivation

$$\begin{aligned} S_0 &\Rightarrow bS_1 \quad &&\Rightarrow bbS_1S_1 \quad &&\Rightarrow bbS_1aS_0 \\ &\Rightarrow bbS_1aaS_2 \quad &&\Rightarrow bbaaaS_2 \quad &&\Rightarrow bbaaab \end{aligned}$$

is not leftmost since in deriving bbS_1aaS_2 from bbS_1aS_0 we do not replace the leftmost nonterminal S_1.

(Example cont'd)

We can transform this derivation into a leftmost derivation by changing the order in which nonterminals are replaced. Namely, in grammar G, we have the leftmost derivation

$$\begin{aligned} S_0 &\Rightarrow bS_1 \Rightarrow bbS_1S_1 \Rightarrow bbaS_1 \\ &\Rightarrow bbaaS_0 \Rightarrow bbaaaS_2 \Rightarrow bbaaab. \end{aligned}$$

Theorem

Let $G = (A_N, A_T, S, P)$ be a context-free grammar. For every complete derivation d of length n in G, $X \Rightarrow \gamma_1 \Rightarrow \cdots \Rightarrow \gamma_n$, where $\gamma_n = u \in A_T^*$, there is a complete leftmost derivation of length n, using the same productions as d, that allows us to derive γ_n from X.

Proof

The argument is by strong induction on $n \geq 1$ for leftmost derivations.
For $n = 1$, the statement is trivially true, since any derivation $X \Rightarrow w_1$ is a leftmost derivation.
Suppose that the statement holds for derivations whose length is no more than n, and let d

$$X \Rightarrow \gamma_1 \Rightarrow \cdots \Rightarrow \gamma_{n+1}$$

be a derivation of length $n+1$. If the first production used in this derivation is $X \to w_0 X_{i_1} w_1 \cdots X_{i_k} w_k$, where $w_i \in A_T^*$ for $0 \leq i \leq k$, then we can write $\gamma_{n+1} = w_0 u_1 w_1 \cdots u_k w_k$, where d_j is a complete derivation $X_{i_j} \overset{*}{\underset{G}{\Rightarrow}} u_j$ of length no greater than n, for $1 \leq j \leq k$.

(Proof cont'd)

By the inductive hypothesis, for each of these derivations d_j, we obtain the existence of the leftmost derivation d_j': $X_{i_j} \underset{G,\text{left}}{\overset{*}{\Rightarrow}} u_j$ for $1 \leq j \leq k$, which uses the same set of productions as d_j. Now, we obtain the existence of the leftmost derivation d':

$$\begin{aligned}
X &\Rightarrow w_0 X_{i_1} w_1 X_{i_2} \ldots X_{i_k} w_k \\
&\overset{*}{\Rightarrow} w_0 u_1 w_1 X_{i_2} \ldots X_{i_k} w_k \text{ (using derivation } d_1') \\
&\overset{*}{\Rightarrow} w_0 u_1 w_1 u_2 \ldots X_{i_k} w_k \text{ (using derivation } d_2') \\
&\vdots \\
&\overset{*}{\Rightarrow} w_0 u_1 w_1 u_2 \ldots u_k w_k \text{ (using derivation } d_k'),
\end{aligned}$$

which concludes our argument.

The Theorem may fail if the derivation is not complete, that is, the final word is not in A_T^*.

Example

Let

$$G = (\{S, X, Y, U, V\}, \{a, b\}, S, \{S \to XY, Y \to UV, X \to a, U \to b, V \to b\})$$

be a context-free grammar. Consider the derivation

$$S \Rightarrow XY \Rightarrow XUV$$

This derivation is not leftmost, and there is no leftmost derivation in G such that $S \overset{*}{\underset{G}{\Rightarrow}} XUV$.

Corollary

Let $G = (A_N, A_T, S, P)$ be a context-free grammar. For every complete derivation d of length n in G, $\gamma_0 \Rightarrow \gamma_1 \Rightarrow \cdots \Rightarrow \gamma_n$, where $\gamma_0 \in (A_N \cup A_T)^+$ and $\gamma_n \in A_T^*$, there is a complete leftmost derivation of length n, using the same productions as d, that allows us to derive γ_n from γ_0.

Proof

Suppose that $\gamma_0 = s_0 \dots s_{k-1}$, where $s_i \in A_N \cup A_T$ for $0 \leq i \leq k-1$. By Theorem 1 we can write $\gamma_n = u_0 \cdots u_{k-1}$ such that $s_i \overset{*}{\underset{G}{\Rightarrow}} u_i \in A_T^*$ for $0 \leq i \leq k-1$. Thus, we obtain the existence of the leftmost derivations $s_i \overset{*}{\underset{G,left}{\Rightarrow}} u_i$ for $0 \leq i \leq k-1$ that use the same productions as the corresponding previous derivations. Starting from these derivations we obtain the leftmost derivation:

$$
\begin{aligned}
\gamma_0 = s_0 s_1 \cdots s_{k-1} & \\
\overset{*}{\underset{G,left}{\Rightarrow}} & \quad u_0 s_1 \cdots s_{k-1} \\
\overset{*}{\underset{G,left}{\Rightarrow}} & \quad u_0 u_1 \cdots s_{k-1} \\
\vdots & \\
\overset{*}{\underset{G,left}{\Rightarrow}} & \quad u_0 u_1 \cdots u_{k-1} = \gamma_n.
\end{aligned}
$$

Definition

A context-free grammar $G = (A_N, A_T, S, P)$ is *ambiguous* if there exists a word $w \in A_T^*$ such that there are at least two leftmost derivations from S to w in G. Otherwise, G is *unambiguous*.

A context-free language can be generated by both ambiguous and unambiguous grammars.

Example

Consider the context-free grammars

$$G_1 = (\{S\}, \{a\}, S, \{S \to SS, S \to a\})$$

and

$$G_2 = (\{S\}, \{a\}, S, \{S \to aS, S \to a\}).$$

They both generate the language $\{a^n \mid n \geq 1\}$.

(Example cont'd)

They both generate the language $\{a^n \mid n \geq 1\}$. Note that in G_1 we have distinct leftmost derivations:

$$\begin{aligned} S &\underset{G_1}{\Rightarrow} SS \underset{G_1}{\Rightarrow} SSS \underset{G_1}{\Rightarrow} aSS \\ &\underset{G_1}{\Rightarrow} aaS \underset{G_1}{\Rightarrow} aaa \end{aligned}$$

and

$$\begin{aligned} S &\underset{G_1}{\Rightarrow} SS \underset{G_1}{\Rightarrow} aS \underset{G_1}{\Rightarrow} aSS \\ &\underset{G_1}{\Rightarrow} aaS \underset{G_1}{\Rightarrow} aaa. \end{aligned}$$

Thus, G_1 is an ambiguous grammar.

(Example cont'd)

On other hand, the equivalent grammar G_2 is unambiguous, since for every a^n, $n \leq 1$, we have exactly one derivation:

$$S \underset{G_2}{\Rightarrow} aS \underset{G_2}{\Rightarrow} a^2 S \cdots \underset{G_2}{\Rightarrow} a^n.$$

Since a language may have both an ambiguous and an unambiguous grammar, it may not be sufficient to examine one grammar to determine whether or not a language is ambiguous.

Definition

Let L be a context-free language. L is *unambiguous* if there is an unambiguous context-free grammar G such that $L = L(G)$.
L is *inherently ambiguous* if every context-free grammar G such that $L(G) = L$ is ambiguous.

The language $\{a^n \mid n \geq 1\}$ is unambiguous.