Overview

- context-free grammars
- context-free languages
- pushdown automata

Context-Free Grammars

This is G_1:

- $A \to 0A1$
- $A \to B$
- $B \to$

Terminology:

- Each line is a *substitution rule* or *production*.
- Each rule has the form: *symbol* $\to$ *string*. The left-hand symbol is a *variable* (usually upper-case).
- A string consists of variables and *terminals*.
- One variable is the *start variable*.

Rules

- Write down the start variable.
- Pick a variable written down and a derivation that starts with that variable.
- Replace that variable with right-hand side of that derivation.
- Repeat until no variables remain.

Example

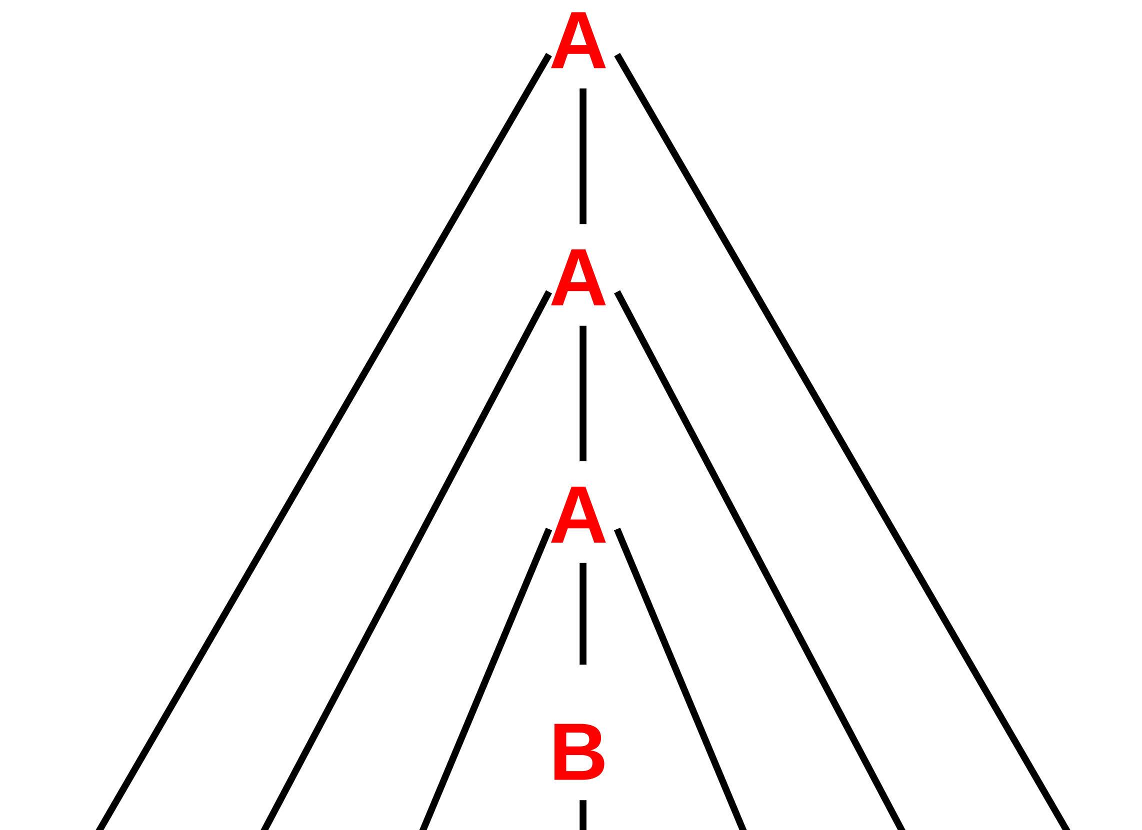
Grammar G_1:

- $A \rightarrow 0A1$
- $A \rightarrow B$
- $B \rightarrow \#$

Derivation with G_1:

$$\begin{aligned} A &\Rightarrow 0A1 \\ &\Rightarrow 00A11 \\ &\Rightarrow 000A111 \\ &\Rightarrow 000B111 \\ &\Rightarrow 000\#111 \end{aligned}$$

A Parse Tree



Context-Free Languages

The language generated in this way is the *language of the grammar*.

For example, $L(G_1)$ is $\{0^n\#1^n | n \geq 0\}$.

Any language generated by a context-free grammar is a *context-free language*.

A Useful Abbreviation

Rules

$$
\begin{aligned}
A &\rightarrow 0A1 \\
A &\rightarrow B
\end{aligned}
$$

are written as:

$$A \rightarrow 0A1 \mid B$$

Example

Grammar:

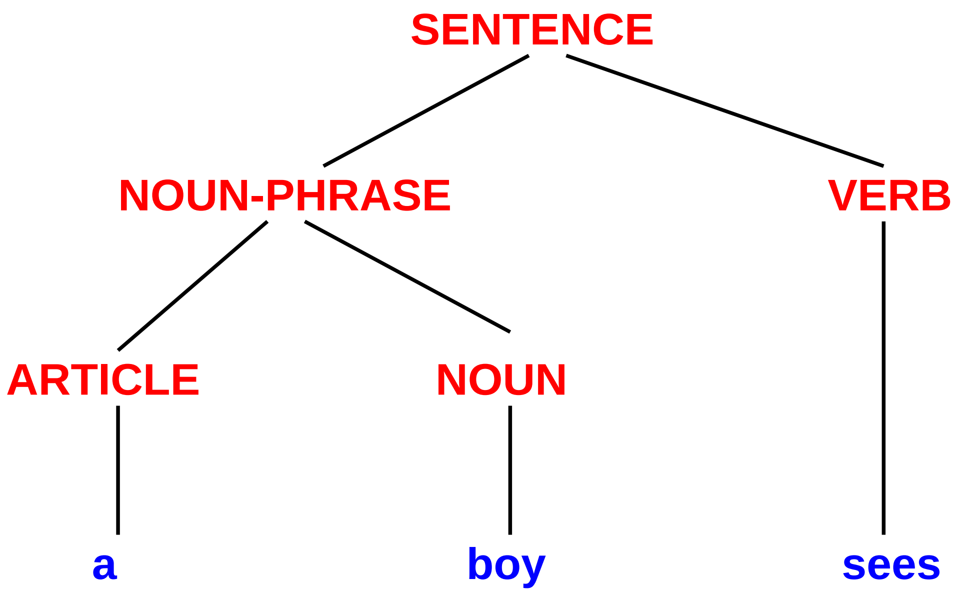
$$
\begin{aligned}
< \text{SENTENCE} > &\rightarrow\ < \text{NOUN-PHRASE} >< \text{VERB} > \\
< \text{NOUN-PHRASE} > &\rightarrow\ < \text{ARTICLE} >< \text{NOUN} > \\
< \text{NOUN} > &\rightarrow\ \text{boy} \mid \text{girl} \mid \text{flower} \\
< \text{ARTICLE} > &\rightarrow\ \text{a} \mid \text{the} \\
< \text{VERB} > &\rightarrow\ \text{touches} \mid \text{likes} \mid \text{sees}
\end{aligned}
$$

Derivation:

$$
\begin{aligned}
< \text{SENTENCE} > &\Rightarrow\ < \text{NOUN-PHRASE} >< \text{VERB} > \\
&\Rightarrow\ < \text{ARTICLE} >< \text{NOUN} >< \text{VERB} > \\
&\Rightarrow\ \text{a}\ < \text{NOUN} >< \text{VERB} > \\
&\Rightarrow\ \text{a boy}\ < \text{VERB} > \\
&\Rightarrow\ \text{a boy sees}
\end{aligned}
$$

Parse Tree

$$
\begin{aligned}
< \text{SENTENCE} > \quad &\Rightarrow \quad < \text{NOUN-PHRASE} >< \text{VERB} > \\
&\Rightarrow \quad < \text{ARTICLE} >< \text{NOUN} >< \text{VERB} > \\
&\Rightarrow \quad \text{a} \;\; < \text{NOUN} >< \text{VERB} > \\
&\Rightarrow \quad \text{a boy} \;\; < \text{VERB} > \\
&\Rightarrow \quad \text{a boy sees}
\end{aligned}
$$



Formal Definitions

A *context-free grammar* is a 4-tuple (V, Σ, R, S) where

- V is a finite set of *variables*,
- Σ is a finite set of *terminals*,
- R is a finite set of *rules*: each rule is a variable and a finite string of variables and terminals.
- S is the *start symbol*.

Formal Definitions

If

- u and v are strings of variables and terminals, and
- $A \to w$ is a rule of the grammar,
- then uAv *yields* uwv, written $uAv \Rightarrow uwv$.

Write $u \stackrel{*}{\Rightarrow} v$ if $u = v$ or

$$u \Rightarrow u_1 \Rightarrow \ldots \Rightarrow u_k \Rightarrow v.$$

for some sequence $u_1, u_2, \ldots, u_k$.

The *language of the grammar* is

$$\left\{ w \mid S \stackrel{*}{\Rightarrow} w \right\}.$$

Example

Consider $G_4 = (V, \Sigma, \{a, b\}, R, S)$.

Rules:

$$S \to aSb \mid SS \mid \varepsilon.$$

Examples: $aaabb$, $aababb$.

Question: What is this language?

Answer: Hint: think of parentheses.

Example

Consider $(V, \Sigma, \{a, b\}, R, E)$ where

- $V = \{E, T, F\}$
- $\Sigma = \{a, +, \times, (,)\}$

Rules:

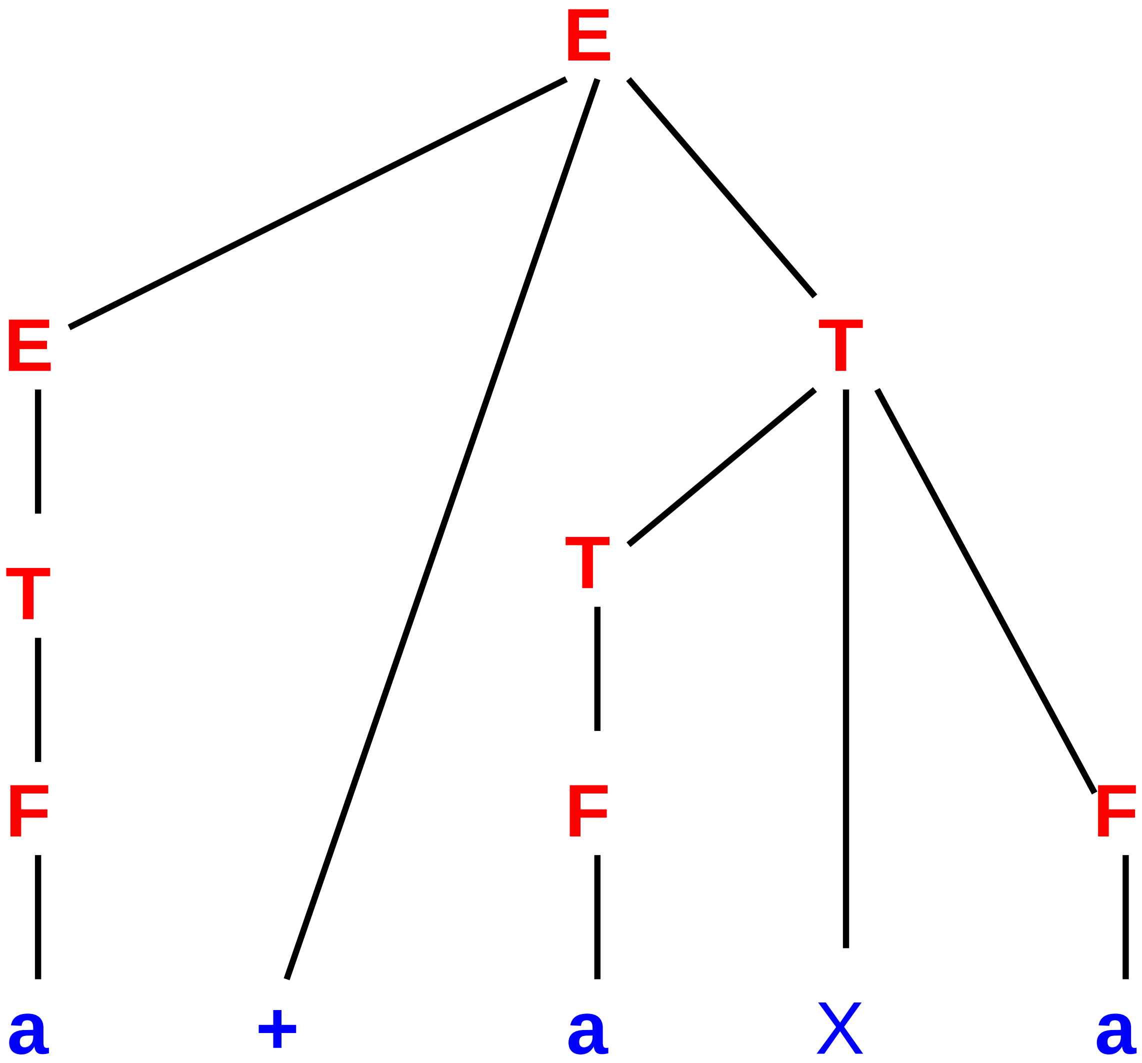
$$\begin{array}{rcl} E & \rightarrow & E \\ T & \rightarrow & T \times F | F \\ F & \rightarrow & (E) | a \end{array}$$

Examples: $a + a \times a$ and $(a + a) \times a$.

Hint: arithmetic expressions. $E =$ expression, $T =$ term, $F =$ factor.

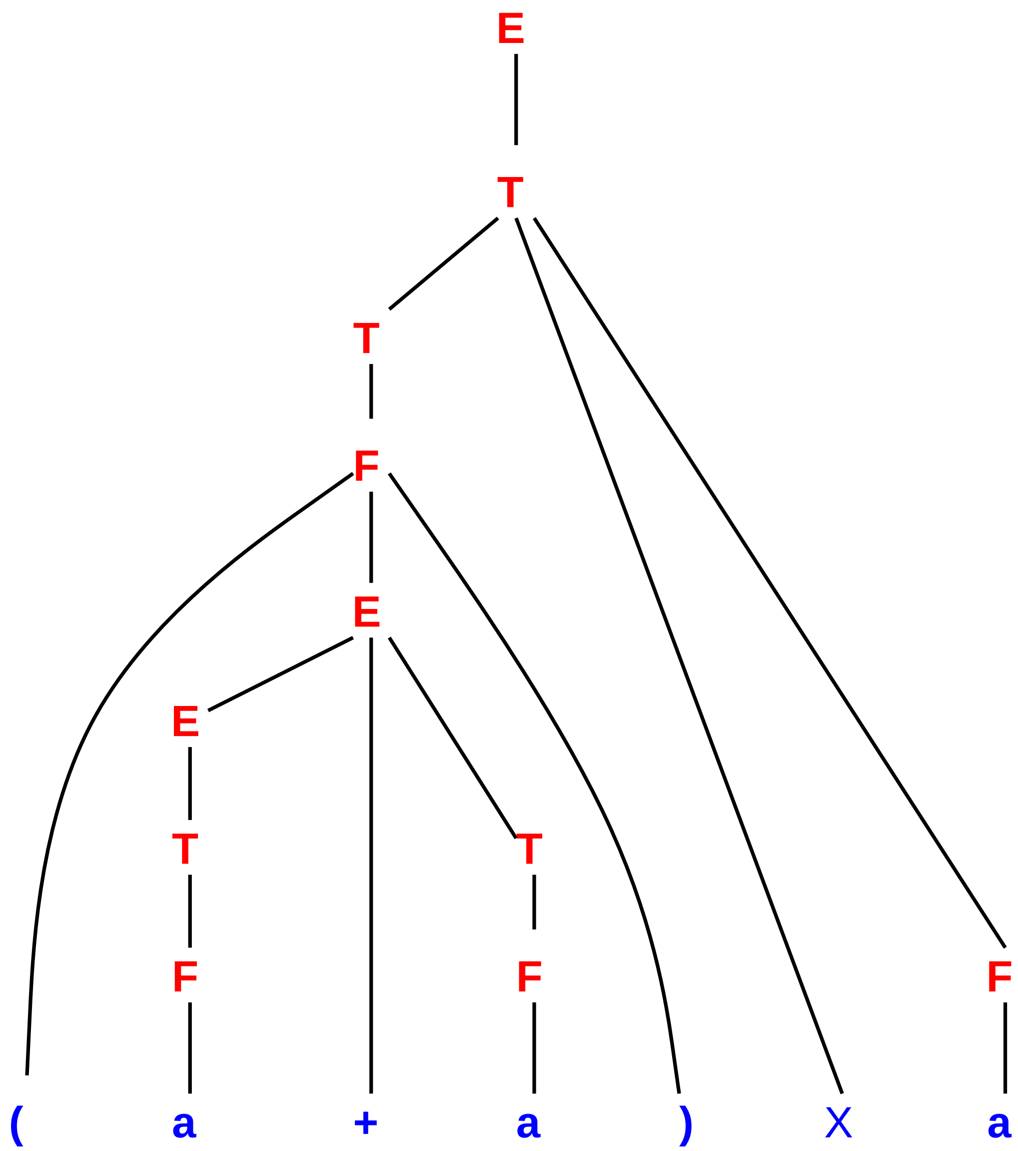
Example

$$
\begin{array}{rcl}
E & \rightarrow & E \\
T & \rightarrow & T \times F | F \\
F & \rightarrow & (E) | a
\end{array}
$$



Example

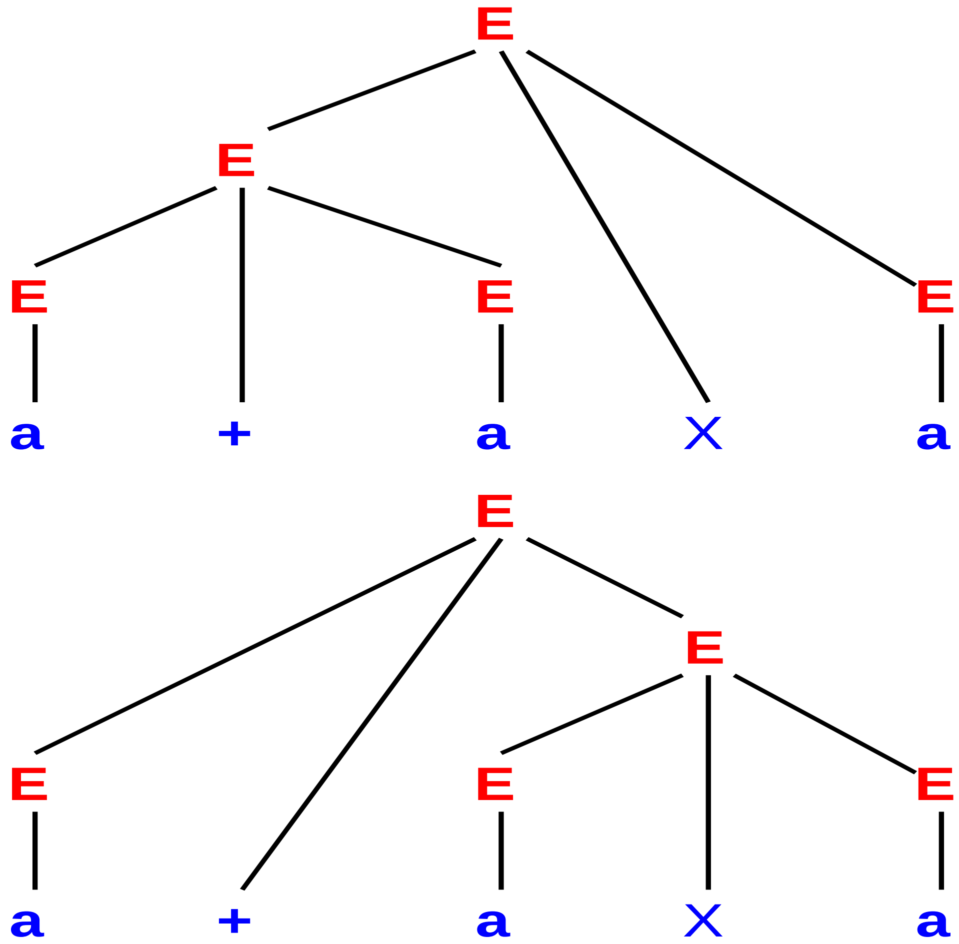
$$
\begin{array}{rcl}
E & \rightarrow & E \\
T & \rightarrow & T \times F | F \\
F & \rightarrow & (E) | a
\end{array}
$$



Ambiguity

Grammar:

$$E \to E + E \mid E \times E \mid (E) \mid a$$



Chomsky Normal Form

Every rule has the form

$$\begin{aligned} A &\rightarrow BC \\ A &\rightarrow a \\ S &\rightarrow \varepsilon \end{aligned}$$

where S is the start symbol, A, B and C are any variable, except B and C not the start symbol.

Theorem

Theorem: Any context-free language is generated by a context-free grammar in Chomsky normal form.

Basic idea:

- Add new start symbol S_0.
- Eliminate all ε rules of the form $A \to \varepsilon$.
- Eliminate all "unit" rules of the form $A \to B$.
- patch up rules so that grammar denotes the same language
- convert remaining rules to proper form.

Proof

Add new start symbol S_0 and rule $S_0 \rightarrow S$.

Guarantees that new start symbol does not appear on right-hand-side of a rule.

Proof

Eliminate ϵ rules.

Repeat:

- remove some $A \to \varepsilon$.
- for each $R \to uAv$, add rule $R \to uv$.
- and so on: for $R \to uAvAw$ add $R \to uvAw$, $R \to uAvw$, and $R \to uvw$.
- for $R \to A$ add $R \to \varepsilon$, except if $R \to \epsilon$ has already been removed.

until all ε-rules not involving the start variable have been removed.

Proof

Eliminate unit rules.

Repeat:

- remove some $A \rightarrow B$.
- for each $B \rightarrow u$, add rule $A \rightarrow u$, unless this is previously removed unit rule.

until all unit rules have been removed.

Proof

Finally, convert remaining rules.

To replace each

$$A \to u_1 u_2 \ldots u_k$$

for $k \geq 3$, introduce new non-terminals

$$N_1, N_2, \ldots, N_{k-1}$$

and rules

$$\begin{array}{rcl} A & \to & u_1 N_1 \\ N_1 & \to & u_2 N_2 \\ & \vdots & \\ N_{k-3} & \to & u_{k-2} N_{k-2} \\ N_{k-2} & \to & u_{k-1} u_k \end{array}$$

Example

Start with:

$$\begin{aligned} S &\rightarrow ASA \mid aB \\ A &\rightarrow B \mid S \\ B &\rightarrow b \mid \varepsilon \end{aligned}$$

Add new start state:

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow ASA \mid aB \\ A &\rightarrow B \mid S \\ B &\rightarrow b \mid \varepsilon \end{aligned}$$

Example

$$\begin{array}{rcl} S_0 & \to & S \\ S & \to & ASA \mid aB \\ A & \to & B \mid S \\ B & \to & b \mid \varepsilon \end{array}$$

Remove ε-rule $B \to \varepsilon$:

$$\begin{array}{rcl} S_0 & \to & S \\ S & \to & ASA \mid aB \mid a \\ A & \to & B \mid S \mid \varepsilon \\ B & \to & b \mid \varepsilon \end{array}$$

Example

$$
\begin{aligned}
S_0 &\rightarrow S \\
S &\rightarrow ASA \mid aB \mid a \\
A &\rightarrow B \mid S \mid \varepsilon \\
B &\rightarrow b
\end{aligned}
$$

Remove ε-rule $A \rightarrow \varepsilon$:

$$
\begin{aligned}
S_0 &\rightarrow S \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \mid S \\
A &\rightarrow B \mid S \mid \varepsilon \\
B &\rightarrow b
\end{aligned}
$$

Example

$$
\begin{array}{rcl}
S_0 & \rightarrow & S \\
S & \rightarrow & ASA \mid aB \mid a \mid AS \mid SA \mid S \\
A & \rightarrow & B \mid S \\
B & \rightarrow & b
\end{array}
$$

Remove unit rule $S \rightarrow S$

$$
\begin{array}{rcl}
S_0 & \rightarrow & S \\
S & \rightarrow & ASA \mid aB \mid a \mid AS \mid SA \mid S \\
A & \rightarrow & B \mid S \\
B & \rightarrow & b
\end{array}
$$

Example

$$
\begin{aligned}
S_0 &\rightarrow S \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow B \mid S \\
B &\rightarrow b
\end{aligned}
$$

Remove unit rule $S_0 \rightarrow S$:

$$
\begin{aligned}
S_0 &\rightarrow S \mid ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow B \mid S \\
B &\rightarrow b
\end{aligned}
$$

Example

$$
\begin{aligned}
S_0 &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow B \mid S \\
B &\rightarrow b
\end{aligned}
$$

Remove unit rule $A \rightarrow B$:

$$
\begin{aligned}
S_0 &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow B \mid S \mid b \\
B &\rightarrow b
\end{aligned}
$$

Example

$$
\begin{aligned}
S_0 &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow S \mid b \\
B &\rightarrow b
\end{aligned}
$$

$$
\begin{aligned}
S_0 &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow S \mid b \mid ASA \mid aB \mid a \mid AS \mid SA \\
B &\rightarrow b
\end{aligned}
$$

Proof

$$
\begin{aligned}
S_0 &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
S &\rightarrow ASA \mid aB \mid a \mid AS \mid SA \\
A &\rightarrow b \mid ASA \mid aB \mid a \mid AS \mid SA \\
B &\rightarrow b
\end{aligned}
$$

Final simplification:

$$
\begin{aligned}
S_0 &\rightarrow AA_1 \mid UB \mid a \mid SA \mid AS \\
S &\rightarrow AA_1 \mid UB \mid a \mid SA \mid AS \\
A &\rightarrow b \mid AA_1 \mid UB \mid a \mid SA \mid AS \\
A_1 &\rightarrow SA \\
U &\rightarrow a \\
B &\rightarrow b
\end{aligned}
$$