Context-free grammars:

$S \in N$ start non-terminal

$$G = (\Sigma, N, S, P)$$

terminals

non-terminals

productions

$P \subseteq N \times (\Sigma \cup N)^*$

finite sets

Context-free grammars:

$$G = (\Sigma, N, S, P)$$

Context-free languages:

$S \to \circ \to \circ \cdots \circ \to w$

$$\boxed{L(G) = \{ w \in \Sigma^* \mid S \to^* w \}}$$

where $\to \ \subseteq (\Sigma \cup N)^* \times (\Sigma \cup N)^*$ consists of all pairs

$w_1 n w_2 \to w_1 u w_2$

for some $(n, u) \in P$
& some w_1, w_2

A context-free grammar for the language $\{a^n b^n \mid n \geq 0\}$

Terminals:

$$a \qquad b$$

Non-terminal:

$$I$$

Start symbol:

$$I$$

Productions:

$$I ::= \varepsilon \mid aIb$$

A context-free grammar for the language $\{a^n b^n \mid n \geq 0\}$

Terminals:

$$a \qquad b$$

Non-terminal:

$$I$$

Start symbol:

$$I$$

Productions:

$$I ::= \varepsilon \mid aIb$$

Any $I \rightarrow^* w$ with $w \in \{a,b\}^*$ has to look like

$I \rightarrow aIb \rightarrow aaIbb \rightarrow \cdots \rightarrow a^n I b^n \rightarrow a^n \varepsilon b^n = a^n b^n$

Some production rules for 'English' sentences

$$
\begin{aligned}
\texttt{SENTENCE} &\to \texttt{SUBJECT VERB OBJECT}\\
\texttt{SUBJECT} &\to \texttt{ARTICLE NOUNPHRASE}\\
\texttt{OBJECT} &\to \texttt{ARTICLE NOUNPHRASE}\\
\texttt{ARTICLE} &\to \texttt{a}\\
\texttt{ARTICLE} &\to \texttt{the}\\
\texttt{NOUNPHRASE} &\to \texttt{NOUN}\\
\texttt{NOUNPHRASE} &\to \texttt{ADJECTIVE NOUN}\\
\texttt{ADJECTIVE} &\to \texttt{big}\\
\texttt{ADJECTIVE} &\to \texttt{small}\\
\texttt{NOUN} &\to \texttt{cat}\\
\texttt{NOUN} &\to \texttt{dog}\\
\texttt{VERB} &\to \texttt{eats}
\end{aligned}
$$

Language generated is regular – it's $L(r)$ for $r =$

$(a|the)(big|small|\varepsilon)(cat|dog)eats(a|the)(big|small|\varepsilon)(cat|dog)$

Every regular language is context-free

Given a DFA M, the set $L(M)$ of strings accepted by M can be generated by the following context-free grammar:

set of terminals = Σ_M

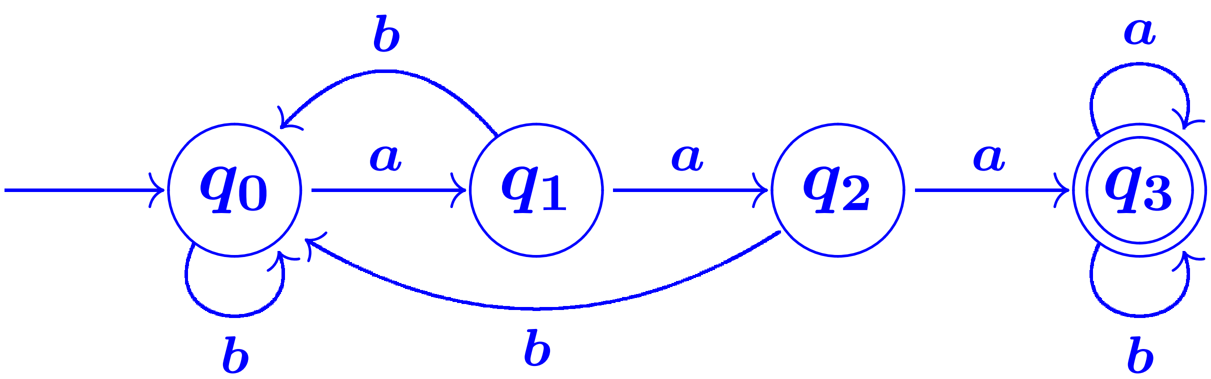
set of non-terminals = $States_M$

start symbol = start state of M

productions of two kinds:

$q \to aq'$	whenever $q \xrightarrow{a} q'$ in M
$q \to \varepsilon$	whenever $q \in Accept_M$

Example of a finite automaton



States: q_0, q_1, q_2, q_3.

Input symbols: a, b.

Transitions: as indicated above.

Start state: q_0.

Accepting state(s): q_3.

Corresponding context-free grammar:

$$q_0 ::= a q_1 \mid b q_0$$
$$q_1 ::= a q_2 \mid b q_0$$
$$q_2 ::= a q_3 \mid b q_0$$
$$q_3 ::= a q_3 \mid b q_3 \mid \varepsilon$$

Definition A context-free grammar is ***regular*** iff all its productions are of the form

$$X \to uY$$

or

$$X \to u$$

where u is a string of terminals and X and Y are non-terminals.

Theorem

(a) *Every language generated by a regular grammar is a regular language (i.e. is the set of strings accepted by some DFA).*

(b) *Every regular language can be generated by a regular grammar.*

aka "right linear"

Definition A context-free grammar is ***regular*** iff all its productions are of the form

$$X \to uY$$

or

$$X \to u$$

where u is a string of terminals and X and Y are non-terminals.

Theorem

(a) *Every language generated by a regular grammar is a regular language (i.e. is the set of strings accepted by some DFA).*

(b) *Every regular language can be generated by a regular grammar.*

[similar result for "left linear" cf grammars]

Example of the construction used in the proof of the Theorem on Slide 40

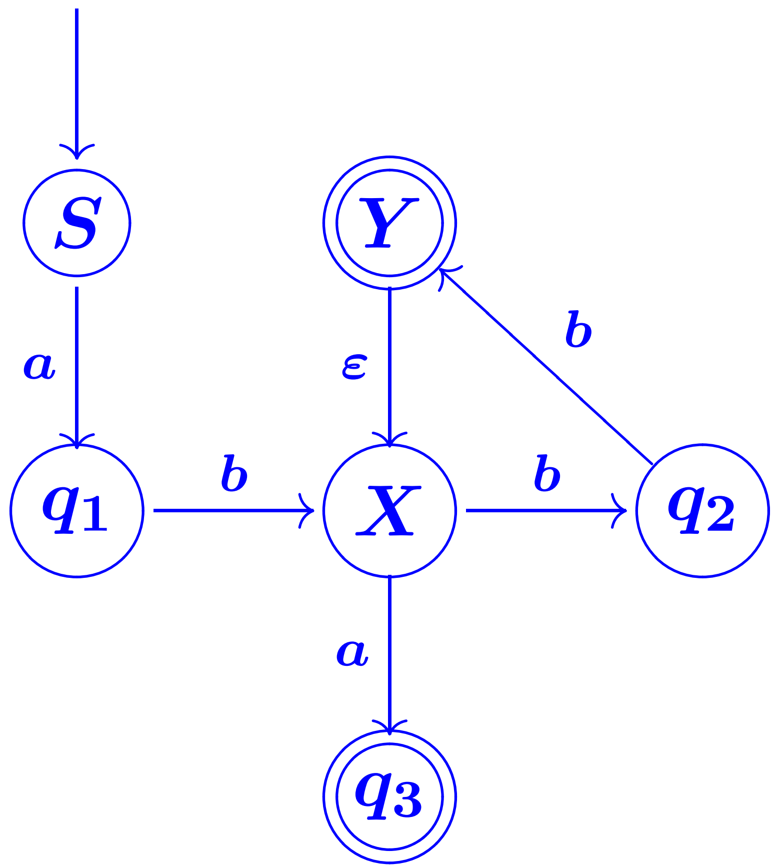
regular grammar:

$$S \to abX$$
$$X \to bbY$$
$$Y \to X$$
$$X \to a$$
$$Y \to \varepsilon$$

(start symbol = S)

$\rightsquigarrow$

NFA$^{\varepsilon}$:



Chomsky Normal Form (CNF)

Theorem

Any context-free language can be generated by a grammar whose productions are of one of the following three types:

$$X \to YZ \qquad X \to a \qquad I \to \varepsilon$$

where X, Y, Z are non-terminals, a is a terminal, and I is the start symbol.

The last type of production occurs if and only if the language contains ε (which is why the use of CNFs is usually restricted to languages that do not contain ε.)

[Example 6.4.1, p 54]

CFG in Chomsky Normal Form for $\{a^n b^n \mid n \geq 0\}$

Terminals : $a \quad b$

Non-terminals : $I \quad A \quad B \quad C$

Start : I

Productions :

$$I ::= \varepsilon \mid AB \mid AC$$

$$A ::= a$$

$$B ::= b$$

$$C ::= IB$$

Greibach Normal Form (GNF)

Theorem

Any context-free language can be generated by a grammar whose productions are of one of the following two types:

$$X \to aU \qquad\qquad I \to \varepsilon$$

where a is a terminal, U is a (possibly empty) string of non-terminals, and I is the start symbol.

The last type of production occurs if and only if the language contains ε (which is why the use of GNFs is usually restricted to languages that do not contain ε.)