

GRAMMARS

A *grammar* is defined (formally) as a 4-tuple $G=(N,\Sigma,P,S)$ where

- N is the set of *nonterminal symbols*.
- Σ is the set of *terminal symbols*.
- P is the set of *production rules*, each of the form $\alpha \rightarrow \beta$.
- S is the *start symbol*.

The *Chomsky hierarchy* defines four distinct types of grammars based on restrictions to the types of production rules allowed in the grammar. These types of grammars are:

1. *Regular grammars*. Productions must be of the form $A \rightarrow xB$, where A and B are nonterminal symbols and x is a string of terminal symbols. For example, the grammar below defining identifiers in a programming language is regular.

$$G = (\{I, X\}, \{a, b, c, \dots, z, 0, 1, \dots, 9\}, P, I)$$

$$I \rightarrow aX \mid bX \mid cX \mid \dots \mid zX \mid a \mid b \mid \dots \mid z$$

$$X \rightarrow aX \mid bX \mid cX \mid \dots \mid zX \mid 0X \mid 1X \mid \dots \mid 9X \mid a \mid b \mid \dots \mid z \mid 0 \mid 1 \mid \dots \mid 9$$

A sample derivation is

$$I \Rightarrow cX \Rightarrow csX \Rightarrow cs4X \Rightarrow cs40X \Rightarrow cs405$$

2. *Context-free grammars*. Productions must be of the form $A \rightarrow \alpha$, where A is a nonterminal symbol and α is a string of terminals and nonterminals in any order. (This is essentially the same as a BNF grammar.) The grammar below is an example of a context-free grammar.

$$G = (\{E, T, F\}, \{a, +, *, (,)\}, P, E)$$

$$E \rightarrow E + T \mid T \qquad T \rightarrow T * F \mid F \qquad F \rightarrow (E) \mid a$$

3. *Context-sensitive grammars*. Productions must be of the form $\alpha \rightarrow \beta$, where α and β are strings on terminals and nonterminals with the restriction that the number of symbols in α must not exceed the number of symbols in β . An example of a context-sensitive grammar is:

$$G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$$

$$S \rightarrow aSBC \mid abC \qquad CB \rightarrow BC$$

$$bB \rightarrow bb \qquad bC \rightarrow bc \qquad cC \rightarrow cc$$

This grammar generates strings of the form $a^n b^n c^n$ where $n > 0$. For example,

$$S \Rightarrow aSBC \Rightarrow aabCBC \Rightarrow aabBCC \Rightarrow aabbCC \Rightarrow aabbcC \Rightarrow aabbcc$$

4. *Unrestricted grammars*. Production rules are of the form $\alpha \rightarrow \beta$ with no restrictions. For example,

$$G = (\{S, A, B, C, D\}, \{a, b\}, P, S)$$

$$S \rightarrow CD \qquad Ab \rightarrow bA$$

$$C \rightarrow aCA \qquad Ba \rightarrow aB$$

$$C \rightarrow bCB \qquad Bb \rightarrow bB$$

$$AD \rightarrow aD \qquad C \rightarrow \epsilon$$

$$BD \rightarrow bD \qquad D \rightarrow \epsilon$$

$$Aa \rightarrow aA$$

This grammar generates symmetrical strings containing a's and b's (e.g. aabaab, baabaa, abbbabbb, etc.). For example,

$$S \Rightarrow CD \Rightarrow aCAD \Rightarrow abCBAD \Rightarrow abBAD \Rightarrow abBaD \Rightarrow abaBD \Rightarrow ababD \Rightarrow abab$$