GRAMMARS

A *grammar* is defined (formally) as a 4-tuple G=(N,Σ,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→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,...,z,0,1,...,9},P,I)
   I → aX | bX | cX | ... | zX | a | b | ... | z
   X → aX | bX | cX | ... | zX | 0X | 1X | ... | 9X | a | b | ... | z | 0 | 1 | ... | 9

   A sample derivation is

   I ⇒ cX ⇒ csX ⇒ cs4X ⇒ cs40X ⇒ cs405

2. *Context-free grammars*. Productions must be of the form A→ α, 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 → E + T | T   T → T * F | F   F → (E) | 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 → aSBC | abC   CB → BC
   bB → bb   bC → bc   cC → cc

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

   S ⇒ aSBC ⇒ aabCBC ⇒ aabBCC ⇒ aabbCC ⇒ aabbcC ⇒ 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 → CD   Ab → bA
   C → aCA   Ba → aB
   C → bCB   Bb → bB
   AD → aD   C → ϵ
   BD → bD   D → ϵ
   Aa → aA

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

   S ⇒ CD ⇒ aCAD ⇒ abCBAD ⇒ abBAD ⇒ abBaD ⇒ abaBD ⇒ ababD ⇒ abab