

Theory of Computation, CSCI 438 spring 2022
Chomsky Normal Form, pg. 108-111, Feb. 18

A context-free grammar is in Chomsky normal form if every rule is of the form

$$A \rightarrow BC$$

$$A \rightarrow \alpha$$

Where α is any terminal and A, B, and C are any variables – except that B and C may not be the start variable. In addition allow the rule $S \rightarrow \epsilon$, where S is the start variable.

(Definition 2.8, page 109)

Theorem 2.9 (page 107) Any context-free language is generated by a context-free grammar in Chomsky normal form.

- Step 1. Add a new start symbol (so we won't have recursion with the start symbol, although of course we may have it afterwards)
- Step 2. Remove ϵ -productions, $A \rightarrow \epsilon$, where A is not the start variable, and updating all productions where A occurs on the right appropriately. Do this over and over until there are no more ϵ -productions.
- Step 3. Remove all unit productions (i.e. a single variable on the right side). We can simply delete productions of the form $X \rightarrow X$. Replace productions of the form $X \rightarrow Y$ with $X \rightarrow$ "whatever Y goes to".
- Step 4. Add new variables as needed to get the grammar to adhere to rule of Chomsky normal form.