CSE396 Lecture Thu. 3/25: Chomsky NF and the CFL Pumping Lemma

Definition: A CFG $G = (V, \Sigma, R, S)$ is in **Chomsky normal form** (**ChNF**) if every rule has the form

- $A \rightarrow c$, with $c \in \Sigma$, or
- $A \rightarrow BC$, where $A, B, C \in V$ (note: we can have $B = C$ or $B, C = A$ etc.)

Most sources also require that the start symbol S cannot be on the right-hand side of a rule. Some (then) allow $S \rightarrow \epsilon$ as the only permitted ϵ-rule. I take the most "liberal" options here.

Theorem: For every CFG G we can build a CFG G' in ChNF such that $L(G') = L(G) \setminus \{\epsilon\}$ (or if we allow the second liberal condition, we get $L(G') = L(G)$ even when $\epsilon \in L(G)$).

We will *skip* the proof for now. The main significance for us is that ChNF makes all parse tree into binary trees.

One other consequence to note later in Chapter 4 as well: If a grammar G in ChNF derives a string x of length $n \geq 1$ at all, then it derives x in exactly $2n-1$ steps, $n-1$ of which use productions of the form $A \rightarrow BC$, and n to fill in terminal symbols one at a time.

[The rest of this lecture was done from the hand-drawn diagrams at
https://cse.buffalo.edu/~regan/cse396/CSE396lect040219.pdf
and the first page of
https://cse.buffalo.edu/~regan/cse396/CSE396lect040419.pdf]