Computer Science 456/656. Context-Sensitive Grammars

A context sensitive grammar, like all the grammars we have introduced in this course consists of the following parts:

1. Σ, the alphabet of *terminals*.
2. V, the alphabet of *variables*.
3. $S \in V$, the *start symbol*.
4. $\mathcal{R}$, a finite set of rules, or productions.

Terminals and variables together are called *grammar symbols*. What makes the grammar context-sensitive is the set of restictions on the productions. You might find different definitions in various places.

Each production has a left-hand side and a right-hand side. Each production is of the form:

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

where A is a variable, α, β, and γ are strings of grammar symbols, and $\gamma \neq \varepsilon$, except as stated below.

A grammar with only productions of the form described above cannot generate the empty string, and thus not every context-free language would be a context-sensitive language. In order to prevent this, we allow one more production, namely $S \rightarrow \varepsilon$, but we make the additional restriction that S cannot appear on the right hand side of any production. With this slight "tweak" to the definition, every context-free language is context-sensitive.

Example

The language $\{a^n b^n c^n\}$ is generated by the following context-sensitive grammar (modified from a Wikipedia page).

1. $S \rightarrow \varepsilon$
2. $S \rightarrow T$
3. $S \rightarrow aTBC$
4. $T \rightarrow aTBC$
5. $S \rightarrow aBC$
6. $T \rightarrow aBC$
7. $CB \rightarrow HB$
8. $HB \rightarrow HC$
9. $HC \rightarrow BC$
10. $aB \rightarrow ab$
11. $bB \rightarrow bb$
12. $bC \rightarrow bc$
13. $cC \rightarrow cc$