1 The Chomsky Hierarchy

1.1 Type-0: Unrestricted Grammars

Unrestricted Grammars are grammars that generate exactly all languages that can be recognized by a turing machine. The language produced by these grammars is called recursively enumerable or Turing-recognizable languages.

Unrestricted grammars have no restrictions to the left and right side of the grammar's productions. This is the most general class of grammars. $G = (N, \Sigma, P, S)$. With N is a set of non-terminal symbols, Σ is a set of terminal symbols, P is a set of production rules of the form $\alpha \to \beta$, where α and β are strings of symbols, $(\alpha\beta) \in (N \cup \Sigma)$ and α is not the empty string, and $S \in N$ is a specially designated start symbol. There are no restriction to the production rules as there are in other grammars. There is a method to prove that any unrestriced grammar ce be represented by a turing machine, and the language a turing machine recognizes can be generated by an unrestriced grammar.

The decision problem of weither a given string can be generated by an unrestriced grammar is equivilent to the halting problem, and so is undecidable.

$$G = (N, \Sigma, P, S)$$

$$x \to y$$

$$x \in (N \cup \Sigma)^+$$

$$y \in (N \cup \Sigma)^*$$

1.2 Type-1: Context-Sensitive Grammars

Context-Sensitive Grammars are grammars that generate context-sensitive languages. Languages produced by these grammars are exactly all languages that can be recognized by a linear bound automaton.

Context-Sensitve Grammars are grammars that can have context of the string in both sides of the production rule. $G = (N, \Sigma, P, S)$. With N is a set of non-terminal symbols,

Σ is a set of terminal symbols, P is a set of production rules with the form $\alpha A\beta \to \alpha\gamma\beta$. Where $A \in N$, $\alpha, \beta \in (N \cup \Sigma)^*$, and $\gamma \in (N \cup \Sigma)^+$.

Left/Right Context-Sensitive Grammars are when the production rules are restricted to the form of $\alpha A \to \alpha\gamma$ or $A\beta \to \gamma\beta$ respectivly. These are the same as normal context-sensitive grammars, with just either α or β representing σ. The Decision problem of weather some string is an element of a context-sensitive language is PSPACE-complete. It has been shown that almost all natural languages can be characterized by a context-sensitive grammar.

$$G = (N, \Sigma, P, S)$$

$$x \to y$$

$$x, y \in (N \cup \Sigma)^+$$

$$|x| \leq |y|$$

$$xAy \to xuy$$

Example:

$$L = a^n b^n c^n : n \geq 1$$

Production Rules:

$$[1] S \rightarrow abc/aAbc$$

$$[2] Ab \rightarrow bA$$

$$[3] Ac \rightarrow Bbcc$$

$$[4] bB \rightarrow Bb$$

$$[5] aB \rightarrow aa/aaA$$

Generating string: $a^3b^3c^3$

$$
\begin{aligned}
S &\Rightarrow aAbc \; [1] \\
&\Rightarrow abAc \; [2] \\
&\Rightarrow abBbcc \; [3] \\
&\Rightarrow aBbbcc \; [4] \\
&\Rightarrow aaAbbcc \; [5] \\
&\Rightarrow aabAbcc \; [2] \\
&\Rightarrow aabbAcc \; [2] \\
&\Rightarrow aabbBbccc \; [3] \\
&\Rightarrow aabBbbccc \; [4] \\
&\Rightarrow aaBbbbccc \; [4] \\
&\Rightarrow aaabbbccc \; [5]
\end{aligned}
$$

References

[1] Hopcroft John, Motwani Rajeev, and Ullman Jeffrey. *Introduction to Automata Theory, Languages, and Computation*. 2nd ed. 2001.

[2] Davis Martin, Sigal Ron, and Weyuker Elane. *Computability, Complexity, and Languages. Fundamentals of Theoretical Computer Science*. 2nd ed. 1994.

[3] Linz Peter. *An Introduction to Formal Languages and Automata*. 5th ed. 2012.