The Chomsky hierarchy

Arno Pauly

February 5, 2021

The basic idea

Grammars with simpler rules are easier to handle.

The hierarchy, overview

From simplest to most complicated:

3. Regular languages (having right-linear grammars)
2. Context-free languages (having context-free grammars)
1. Context-sensitive languages (having context-sensitive grammars)
0. Computably enumerable languages (having (unrestricted) grammars)
-1. Arbitrary languages (not necessarily describable by a grammar at all)

Right-linear grammars

Definition

A grammar is *right-linear*, if all rules are of the form $T \to \varepsilon$ or $T \to aR$ for $T, R \in \mathcal{N}$ and $a \in \Sigma$.

We shall also allow the form $T \to a$ as abbreviation for $T \to aQ$, $Q \to \varepsilon$ for a fresh non-terminal Q.

Definition

A grammar is *left-linear*, if all rules are of the form $T \to \varepsilon$ or $T \to Ra$ for $T, R \in \mathcal{N}$ and $a \in \Sigma$.

Theorem

Right-linear and left-linear grammars describe the same languages.

Context-free grammars

Definition

A grammar is *context-free*, if the left-hand side of every rule is a single non-terminal.

Context-sensitive grammars

Definition

A grammar is *context-sensitive*, if every rule is of the form $wAu \to wvu$ where $v \neq \varepsilon$; or is $S \to \varepsilon$, and where S never appears on the right-hand side of a rule.

Definition

A grammar is *monotonic*, if for all rules $u \to w$ (except potentially $S \to \varepsilon$) it holds that $|u| \leq |w|$, and S never appears on the right-hand side of a rule.

Theorem

Context-sensitive grammars and monotonic grammars describe the same languages.

Upcoming

Next, we'll have a closer look at regular languages and see how we can exploit their structure, recognize them with finite automata, and so on.