Chomsky Hierarchy

Jiří Techet Tomáš Masopust Alexander Meduna

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, Brno 61266, Czech Republic

Modern Formal Language Theory, 2007

Chomsky Hierarchy

Chomsky Hierarchy

$$\mathscr{L}(REG) \subset \mathscr{L}(LIN) \subset \mathscr{L}(CF) \subset \mathscr{L}(CS) \subset \mathscr{L}(RE) \subset \mathscr{L}(ALL)$$

$\mathscr{L}(REG)$ family of regular languages (type 3)

$\mathscr{L}(LIN)$ family of linear languages

$\mathscr{L}(CF)$ family of context-free languages (type 2)

$\mathscr{L}(CS)$ family of context-sensitive languages (type 1)

$\mathscr{L}(RE)$ family of recursively enumerable languages (type 0)

$\mathscr{L}(ALL)$ family of all languages

Type 0 Grammars

Type 0 Grammar

$$G = (N, T, P, S)$$

N alphabet of nonterminals

T alphabet of terminals

P finite set of productions (rules) of the form

$$y \to x$$

with $y, x \in V^*$, $\text{alph}(y) \cap N \neq \emptyset$

S the start symbol, $S \in N$

- $V = N \cup T$
- $\text{alph}(y)$ is the set of all symbols occurring in $y \in V^*$

Derivation Step

Production Label

Consider a production $p : y \to x$

p label

$y \to x$ production

Instead of $y \to x \in P$, we often write $p \in P$

Derivation Step

For every $u, v \in V^*$ and $p : y \to x$,

$$uyv \Rightarrow uxv \; [p]$$

or simply

$$uyv \Rightarrow uxv$$

Notation

Notation

For every $y_0, y_1, \ldots, y_n$, for some $n \geq 1$, such that

$$y_0 \Rightarrow y_1 \, [p_1] \Rightarrow y_2 \, [p_2] \Rightarrow \ \ldots \ \Rightarrow y_n \, [p_n],$$

where $p_i \in P$, for all $i = 1, \ldots, n$,

- $y_0 \Rightarrow^n y_n \, [p_1 \ldots p_n]$ or $y_0 \Rightarrow^n y_n$
- for every $y \in V^*$,
$$y \Rightarrow^0 y \, [\varepsilon] \text{ or } y \Rightarrow^0 y$$
- if $v \Rightarrow^m w \, [\alpha]$ for some $m \geq 1$, then
$$v \Rightarrow^+ w \, [\alpha] \text{ or } v \Rightarrow^+ w$$
- if $v \Rightarrow^m w \, [\alpha]$ for some $m \geq 0$, then
$$v \Rightarrow^* w \, [\alpha] \text{ or } v \Rightarrow^* w$$

Language of Grammar

Relation of Direct Derivation

$\Rightarrow$ direct derivation relation

$\Rightarrow^*$ transitive and reflexive closure of direct derivation relation

$\Rightarrow^+$ transitive closure of direct derivation relation

Language of Grammar

$$L(G) = \{w \in T^* : S \Rightarrow^* w\}$$

CS, *CF*, *LIN* and *REG* Grammars

CS, *CF*, *LIN* and *REG* Grammars

A type-0 grammar $G = (N, T, P, S)$ is

1. context-sensitive if for every $y \to x \in P$,

$$|y| \leq |x|, \text{ or } y = S, x = \varepsilon$$

2. context-free if for every $y \to x \in P$,

$$y \in N$$

3. linear if for every $y \to x \in P$,

$$y \in N, x \in T^* \cup T^*NT^*$$

4. regular if for every $y \to x \in P$,

$$y \in N, x \in \{\varepsilon\} \cup T \cup TN$$

Bibliography


N. Chomsky.
Three models for the description of language.
Transactions on Information Theory, 2(3):113–124, 1956.


A. Meduna.
Automata and Languages: Theory and Applications.
Springer, London, 2000.


G. Rozenberg and A. Salomaa.
Handbook of Formal Languages, volume 1–3.
Springer, Berlin, 1997.


A. Salomaa.
Formal Languages.
Academic Press, New York, 1973.