Chomsky Hierarchy

Type	Grammar	Language	Automata
3	Finite State	Regular	Finite
2	Context-Free	C-F	Pushdown
1	Context-Sensitive	C-S	Linear-Bounded
0	General Rewrite	Unrestricted	Turing Machines

- occasionally referred to as Chomsky–Schützenberger hierarchy
- described by Noam Chomsky in 1956
- hierarchy of grammars
    - Type 3 $\subset$ Type 2 $\subset$ Type 1 $\subset$ Type 0

Alphabets and Languages

An <u>alphabet</u> is a finite non-empty set.

Let S and T be alphabets.

$S \bullet T = \{ s\,t \mid s \in S, t \in T \}$

(We'll often write ST for S•T.)

λ = empty string, string of length one

$S^0 = \{\lambda\}$

$S^1 = S$

$S^n = S^{(n-1)} \bullet S, n > 1$

$S^+ = S^1 \cup S^2 \cup S^3 \cup \ldots$

$S^* = S^0 \cup S^+$

A <u>language</u> L over an alphabet S is a subset of S*.

Definition of a Grammar

A grammar G is a 4 tuple G = (N, Σ, P, S), where

N is an alphabet of nonterminal symbols

Σ is an alphabet of terminal symbols

N and Σ are disjoint

S is an element of N; S is the start symbol or initial symbol of the grammar

P is a set of productions of the form α -> β where

α is in (N U Σ)* N (N U Σ)*

β is in (N U Σ)*

Classes of Grammars (The Chomsky Hierarchy)

Type 0, Phrase Structure (same as basic grammar definition)

Type 1, Context Sensitive

(1) $\alpha \rightarrow \beta$ where α is in $(N \cup \Sigma)^* N (N \cup \Sigma)^*$,
β is in $(N \cup \Sigma)+$, and $\text{length}(\alpha) \leq \text{length}(\beta)$

(2) $\gamma A \delta \rightarrow \gamma \beta \delta$ where A is in N, β is in $(N \cup \Sigma)^+$, and
γ and δ are in $(N \cup \Sigma)^*$

Type 2, Context Free

$A \rightarrow \beta$ where A is in N, β is in $(N \cup \Sigma)^*$

Linear

$A \rightarrow x$ or $A \rightarrow x B y$, where A and B are in N and x and y are in Σ^*

Type 3, Regular Expressions

(1) left linear $A \rightarrow B a$ or $A \rightarrow a$, where A and B are in N and a is in Σ

(2) right linear $A \rightarrow a B$ or $A \rightarrow a$, where A and B are in N and a is in Σ

Comments on the Chomsky Hierarchy (1)

Definitions (1) and (2) for context sensitive are equivalent.

Definitions (1) and (2) for regular expressions are equivalent.

If a grammar has productions of all three of the forms described in definitions (1) and (2) for regular expressions, then it is a linear grammar.

Each definition of context sensitive is a restriction on the definition of phrase structure.

Every context free grammar can be converted to a context sensitive grammar with satisfies definition (2) which generates the same language except the language generated by the context sensitive grammar cannot contain the empty string λ.

The definition of linear grammar is a restriction on the definition of context free.

The definitions of left linear and right linear are restrictions on the definition of linear.

Comments on the Chomsky Hierarchy (2)

- Every language generated by a left linear grammar can be generated by a right linear grammar, and every language generated by a right linear grammar can be generated by a left linear grammar.
- Every language generated by a left linear or right linear grammar can be generated by a linear grammar.
- Every language generated by a linear grammar can be generated by a context free grammar.
- Let L be a language generated by a context free grammar. If L does not contain λ, then L can be generated by a context sensitive grammar. If L contains λ, then L-$\{\lambda\}$ can be generated by a context sensitive grammar.
- Every language generated by a context sensitive grammar can be generated by a phrase structure grammar.

The Chomsky Hierarchy and the Block Diagram of a Compiler