Languages

Definition: An *alphabet* is a finite, nonempty set of symbols. We use Σ to denote this alphabet.

A *string* is a finite sequence of symbols from Σ.

The length of a string s, denoted $|s|$, is the number of symbols in it.

The empty string, ϵ, is the string of length zero.

Σ^* denotes the set of all sequences of strings that are composed of zero or more symbols of Σ.

Σ^+ denotes the set of all sequences of strings composed of one or more symbols of Σ, i.e. $\Sigma^* - \{\epsilon\}$.

A *language* is a subset of Σ^*.

Grammars

Defined as a quadruple $\langle V_t, V_n, P, S \rangle$, where:

- V_t = the *terminal vocabulary*, i.e. the words of the language: finite (but productive ...)

- V_n = the *non-terminals*, i.e. the set of categories, used for the benefit of the grammar to express generalizations over items in V_t

- P = the *set of productions*, i.e. rules in the grammar: finite, and of the form $\alpha \rightarrow \beta$, where $\alpha, \beta \in (V_n \cup V_t)$

- S = the *distinguished symbol*, or 'start' symbol (cf. 'sentence', if we're talking natural languages):
  - $S \in V_n$,
  - P must include at least one rule $\alpha \rightarrow \beta$ where $\alpha = S$.

A grammar is *generative* if it predicts (explicitly defines, or characterizes) *all and only all* strings $\in$ the language.

Chomsky Hierarchy

There are different types of languages, and grammars whose *power* depends on the nature of $\alpha, \beta \in P$.

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

Type 3 $\subset$ Type 2 $\subset$ Type 1 $\subset$ Type 0