$$\Sigma = \{ a, b, c \}$$

alphabet

symbols

string:
a sequence of symbols

Examples of strings over Σ:

abc   abccba   ab ε

A string is **over** Σ if ___all its symbols are in Σ___

A **language** over Σ is ___a set of strings over Σ___

Examples of languages over Σ:

$\{ aa, bb, cc \}$ $\{ abc, abccba, ab, \varepsilon \}$ $\{ a^n \mid n \geq 0 \}$

$a^2 = aa$ $a^3 = aaa$

We can use **regular expressions** to describe these languages.

Regular Expressions

Basis steps:

   a is a regular expression, for $a \in \Sigma$

   ε is a regular expression

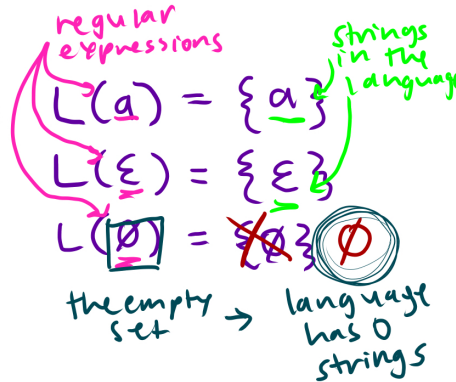 $\emptyset$ is a regular expression

{ }

Recursive steps:

(R_1^*) is a regular expression when R_1 is a regular expression

$(R_1 \circ R_2)$ is a regular expression when R_1 and R_2 are regular expressions

$(R_1 \cup R_2)$ is a regular expression when R_1 and R_2 are regular expressions

regular expressions

$L(a) = \{a\}$ strings in the language

$L(\varepsilon) = \{\varepsilon\}$

$L(\emptyset) = \{\emptyset\}$ $\emptyset$

the empty set → language has 0 strings

$\Sigma = \{a, b, c\}$

Regular Expressions

(R_1^*) is a regular expression when R_1 is a regular expression

b^* $L(b^*) = \{\varepsilon, b, bb, bbb, ...\}$ Σ^* $L(\Sigma^*) =$ the set of all strings over Σ
$\{b^n \mid n \geq 0\}$

$(R_1 \circ R_2)$ is a regular expression when R_1 and R_2 are regular expressions

$a \circ a \circ a \circ a \circ a$ $L(aaaaa) =$ $c \circ b^*$ $L(cb^*) =$
(shorthand: $aaaaa$) $\{aaaaa\}$ (shorthand: cb^*) $\{c, cb, cbb, ...\}$
$\{cb^n \mid n \geq 0\}$

$(R_1 \cup R_2)$ is a regular expression when R_1 and R_2 are regular expressions

$a \cup c$ $L(a \cup c) =$ $cb^* \cup aaaaa$ $L(cb^* \cup aaaaa)$
$L(a) \cup L(c) =$ $cbbaaaaa = $ $= L(cb^*) \cup L(aaaaa)$
$\{a, c\}$ $\{cb^n \mid n \geq 0\} \cup \{aaaaa\}$

$\Sigma = \{a, b, c\}$

Regular Expressions

$$cb^* \cup aaaaa$$

$$L(cb^* \cup aaaaa) = \{c(b)^n \mid n \geq 0\} \cup \{aaaaa\}$$

$$cb \quad cbbbbb$$

Implictly:

$$(c(b^*)) \cup (aaaaa)$$

What happens when we evaluate in a different order?

$$(cb)^* \cup (aaaaa)$$

$$L(cb) = \{cb\}$$

$$L((cb)^*) = \{(cb)^n \mid n \geq 0\}$$

$$\varepsilon \quad cb \quad cbcb \quad cbcbcb \quad \cdots$$

→ The language is not the same!

Precedence order: First *, then ∘, then ∪