Languages

What is a language?

- A language is:
  - A set of strings (finite sequences).
- The language can contain infinitely many strings.
- We need ways in which we can define languages.

Finite Languages

- Simple:
  - We just list the elements in the language.
- For example:
  - {a, ab}
- Given a sequence x we use |x| to denote the length of x, eg:
  - $|aa| = 2$
  - $|a| = 1$
  - $|\varepsilon| = 0$

Compositional Approaches

- Even if a language is finite, it may be very large
  - Example: set of even integers less than 10^{100}

- We don't want to just list elements

- We want ways of combining (composing) strings and languages

Concatenation

- Simply follow one string by the other.
- For example
  - String aa followed by string b is:
  - String aab

- We can raise to power:
  - a^n is a repeated n times.
  - Example: a^2 is aa.

Concatenation: languages

- If A and B are languages then:
  - AB is the set of strings that can be formed by following a string in A by a string in B.
- Example:
  - $\{a,b\}\{aa\} = \{aaa,baa\}$
  - $\{a\}\{aa,b\} = \{aaa,ab\}$
  - $\{\varepsilon\}\{aa,b\} = \{aa,b\}$
  - $\{\varepsilon,a\}\{\varepsilon,a\} = \{\varepsilon,a,aa\}$

Concatenation: properties

- Examples of properties (for any language A):
  - $A\,\{\varepsilon\} = A$
  - $A\,\{\} = \{\}$

- Given strings a, b, c we have:
  - $a(bc) = (ab)c$ (concatenation is associative)
  - $\varepsilon\, a = a = a\, \varepsilon$ (ε is the **unit** of concatenation)

Raising to power

- Given language A, we define AA as expected:
  - The set of strings that can be formed by concatenating two strings from A.
- Examples:
  - $\{a\}^2 = \{aa\}$
  - $\{aa,b\}^2 = \{aaaa,aab,baa,bb\}$
  - $\{a,b,c\}^2 = \{aa,ab,ac,ba,bb,bc,ca,cb,cc\}$

Raising to Power

- We can extend this as expected:
  - $A^0 = \{\varepsilon\}$
  - $A^1 = A$
  - $A^2 = AA$
  - $A^3 = AAA$
  - …
  - $A^n = A^{n-1}A$

  - Example:
    - $\{a,b\}^3 = \{aaa,aab,aba,abb,baa,bab,bba,bbb\}$

Kleene Star

- Given language A, A^* is the union of
  - A^0, A^1, A^2, …
- Given language A we have that A^+ is the union of
  - A^1, A^2, A^3, …
- And so $A^+ = AA^*$ and A^* is the union of $\{\varepsilon\}$ and A^+.
- Also $A^*A^* = A^*$.
- They allow one to build infinite languages

Finally, since languages are sets …

Some properties of sets

- Alphabet Σ
- Basic operators:

$$A \cup B = \{\sigma \in \Sigma^* | \sigma \in A \vee \sigma \in B\}$$

$$A \cap B = \{\sigma \in \Sigma^* | \sigma \in A \wedge \sigma \in B\}$$

$$A \setminus B = \{\sigma \in \Sigma^* | \sigma \in A \wedge \sigma \notin B\}$$

$$\bar{A} = \{\sigma \in \Sigma^* | \sigma \notin A\}$$

- Note, for complement also use:

$$\sim A = \{\sigma \in \Sigma^* | \sigma \notin A\}$$

Some properties of sets

- Associativity

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(AB)C = A(BC)$$

- Commutativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Some properties of sets

- Identity

$$A \cup \{\} = A$$

$$A \cap \Sigma^* = A$$

$$A\{\epsilon\} = A$$

- Zero:

$$A \cap \{\} = \{\}$$

$$A\{\} = \{\}$$

Distributivity

- We have

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A(B \cup C) = AB \cup AC$$
$$(A \cup B)C = AC \cup BC$$

De Morgan Law

Finally

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$