1. Context-Free Grammars

- The **set of sequences** over a set X, called X^* is another set such that:
  - ε is a sequence called the empty sequence, and
  - if z is a sequence and $a \in X$ then az is also a sequence. We use lowercase letters from the end of the latin alphabet to denote sequences: $z, y, x, \ldots$
- An **alphabet** is a finite set X. Its elements are called symbols and are usually denoted by lowercase letters from the beginning of the latin alphabet: $a, b, c, \ldots$
- A **language** is a subset of T^* for some alphabet T. Examples of languages over an alphabet T are $T^*, \emptyset, \{\varepsilon\}$ and T itself.
- A **sentence** (often called **word**) is any element of a language.
- **Language operations.** Let L and M be languages over an alphabet T. Then,
  - $\overline{L} = T^* - L$ is the complement of L,
  - $L^R = \{s^R \mid s \in L\}$ is the reverse of L,
  - $LM = \{st \mid s \in L,\ t \in M\}$ is the concatenation of L and M,
  - $L^0 = \{\varepsilon\}$ is the 0-th power of L,
  - $L^{n+1} = LL^n$ is the $(n+1)$-th power of L,
  - $L^* = \bigcup_{i \in \mathbb{N}} L^i$ is the start-closure of L,
  - $L^+ = \bigcup_{i \in \mathbb{N},\ i>0}$ is the positive clousre of L. And the following properties hold:
  - $L^+ = LL*$, and
  - $L^* = \{\varepsilon\} \cup L^+$.
- A **grammar** is a shorthand syntax for an inductive definition of a language, where we use
  - Monospace characters for **terminals**, i.e. symbols of the alphabet.
  - Capital letters for **nonterminals**, i.e. auxiliary symbols that are not part of the alphabet but are part of the language.
  - **Production rules** of the form $\alpha \to \beta$, where α is always a non-terminal.
  - A **nonterminal start symbol**, which can be ε.
- A **context-free grammar** is a four-tuple (T, N, R, S) where
  - T is a finite set of terminal symbols (alphabet),
  - N is a finite set of nonterminal symbols,
  - R is a finite set of production rules of the form $A \to \beta$, where A is exactly one nonterminal and β is a sequence of terminals and nonterminals, and
  - S is the start symbol.

  An example of a context-free grammar is

  $$P \to \varepsilon \mid a \mid b \mid c \mid aPa \mid bPb \mid cPc.$$

  A non example of a context-free grammar is $\{a^n b^n c^n \mid n \in \mathbb{N}\}$, because in this case the production rules would not be of the form $A \to \beta$ where A is exactly one nonterminal.
- A **context-free language** is a language generated by a context-free grammar.