9. Automata and Monadic Second-Order Logic

Logics on words

- Regular expressions give operational descriptions of regular languages.
- Often the natural description of a language is declarative:
  - **even number of a's and even number of b's** vs.
    $\left(aa + bb + (ab + ba)(aa + bb)^*(ba + ab)\right)^*$
  - **words not containing 'hello'**
- Goal: find a declarative language able to express all the regular languages, and only the regular languages.

Logics on words

- Idea: use a logic that has an interpretation on words
- A formula expresses a property that each word may satisfy or not, like
  - **the word contains only a's**
  - **the word has even length**
  - **between every occurrence of an a and a b there is an occurrence of a c**
- Every formula (indirectly) defines a language: the language of all the words over the given fixed alphabet that satisfy it.

First-order logic on words

- Atomic formulas: for each letter a we introduce the formula $Q_a(x)$, with intuitive meaning: **the letter at position x is an a**.

First-order logic on words: Syntax

- Formulas constructed out of atomic formulas by means of standard “logic machinery”:
  - Alphabet $\Sigma = \{a, b, \ldots\}$ and position variables $V = \{x, y, \ldots\}$
  - $Q_a(x)$ is a formula for every $a \in \Sigma$ and $x \in V$.
  - $x < y$ is a formula for every $x, y \in V$
  - If $\varphi, \varphi_1, \varphi_2$ are formulas then so are $\neg\varphi$ and $\varphi_1 \vee \varphi_2$
  - If φ is a formula then so is $\exists x\ \varphi$ for every $x \in V$

Abbreviations

$$\varphi_1 \wedge \varphi_2 := \neg(\neg\varphi_1 \vee \neg\varphi_2)$$

$$\varphi_1 \rightarrow \varphi_2 := \neg\varphi_1 \vee \varphi_2$$

$$\forall x\, \varphi := \neg\exists\, x \neg\varphi$$

$$\begin{aligned} \mathrm{first}(x) &:= \\ \mathrm{last}(x) &:= \\ y = x + 1 &:= \\ y = x + 2 &:= \\ y = x + (k+1) &:= \end{aligned}$$

Examples (without semantics yet)

- "The last letter is a b and before it there are only a's."

$$\exists x\, Q_b(x) \land \forall x\, (\mathrm{last}(x) \rightarrow Q_b(x) \;\land\; \neg\mathrm{last}(x) \rightarrow Q_a(x))$$

- "Every a is immediately followed by a b."

$$\forall x\, (Q_a(x) \rightarrow \exists y\, (y = x + 1 \land Q_b(y)))$$

- "Every a is immediately followed by a b, unless it is the last letter."

$$\forall x\, (Q_a(x) \rightarrow \forall y\, (y = x + 1 \rightarrow Q_b(y)))$$

- "Between every a and every later b there is a c."

$$\forall x \forall y\, (Q_a(x) \land Q_b(y) \land x < y \rightarrow \exists z\, (x < z \land z < y \land Q_c(z)))$$

First-order logic on words: Semantics

- Formulas are interpreted on pairs $(w, \mathcal{J})$ called interpretations, where
  - w is a word, and
  - $\mathcal{J}$ assigns positions to the free variables of the formula (and maybe to others too—who cares)
- It does not make sense to say a formula is true or false: it can only be true or false for a given interpretation.
- If the formula has no free variables (if it is a sentence), then for each word it is either true or false.

- Satisfaction relation:

$$
\begin{array}{lllll}
(w,\mathfrak{I}) & \models & Q_a(x) & \textit{iff} & w[\mathfrak{I}(x)] = a \\
(w,\mathfrak{I}) & \models & x < y & \textit{iff} & \mathfrak{I}(x) < \mathfrak{I}(y) \\
(w,\mathfrak{I}) & \models & \neg\varphi & \textit{iff} & (w,\mathfrak{I}) \not\models \varphi \\
(w,\mathfrak{I}) & \models & \varphi 1 \vee \varphi_2 & \textit{iff} & (w,\mathfrak{I}) \models \varphi_1 \textit{ or } (w,\mathfrak{I}) \models \varphi_2 \\
(w,\mathfrak{I}) & \models & \exists x\, \varphi & \textit{iff} & |w| \geq 1 \textit{ and some } i \in \{1,\ldots,|w|\} \textit{ satisfies } (w,\mathfrak{I}[i/x]) \models \varphi
\end{array}
$$

- More logic jargon:
  - A formula is valid if it is true for all its interpretations
  - A formula is satisfiable if is is true for at least one of its interpretations

The empty word ...

- ... is as usual a pain in the eh, neck.
- It satisfies all universally quantified formulas, and no existentially quantified formula.

Can we only express regular languages? Can we express all regular languages?

- The language $L(\varphi)$ of a sentence φ is the set of words that satisfy φ.
- A language L is expressible in first-order logic or FO-definable if some sentence φ satisfies $L(\varphi) = L$.
- Proposition: a language over a one-letter alphabet is expressible in first-order logic iff it is finite or co-finite (its complement is finite).
- Consequence: we can only express regular languages, but not all, not even the language of words of even length.