2.4 Ehrenfeucht-Fraïssé Games

For LTL the simple observation that we can describe the model-checking problem by a finite automaton immediately lead to a result that certain properties are not expressible by LTL-formulae. To show get concrete examples of properties not expressible in first-order logic, we need slightly more elaborate techniques.

So, what does "inexpressibility" mean? To show that a certain property (like "the universe has even size") *cannot* be expressed by a certain class $\mathcal{L}$ of formulae, it certainly suffices to exhibit two structures $\mathcal{A}$ and $\mathcal{B}$ such that

> $\mathcal{A}$ has the property, $\mathcal{B}$ doesn't
>
> For all $\varphi \in \mathcal{L}$
>
> $$\mathcal{A} \models \varphi \quad \text{iff} \quad \mathcal{B} \models \varphi$$

So, the problem we're looking at is how to construct different structures that are hard to distinguish. Of course, the smaller $\mathcal{L}$ is the easier it is to come up with examples. Since we will need the last property several times, we introduce a notation for it.

> **Definition.** $\mathcal{A}, \vec{a} \equiv_{\mathcal{L}} \mathcal{B}, \vec{b}$ iff for all $\varphi \in \mathcal{L}$, either both $\mathcal{A}, \vec{a} \models \varphi$ and $\mathcal{B}, \vec{b} \models \varphi$ or neither.
>
> Here $\vec{a} = a_1, \ldots, a_\ell$ and $\vec{b} = b_1, \ldots, b_\ell$.

So, how can first order structures look different? Let's look at some examples.

Words—Spot the difference!

`a a b a a` *vs* `a a a a a`

`a a b a a` *vs* `a a a a b`

`a b a c a` *vs* `a a b c a`

In the first example, the left-hand structure has a letter `b`, which the right-hand structure doesn't.

In the second example, both structures have a letter `b`, but on the left-hand structure we can find an `to` the right of it.

In the third example both structures have a `b` and the same set of letters (`a`, `c`) to the right of it. But only in the left-hand structure we can find a `b` and a `c` with an `a` between them.

Generalising this sequence of more and more complicated differences between structures, we define a game that mimics this step by step discovery of a difference in the structure.

The game, called Ehrenfeucht-Fraïssé game, is a game for two players.

- **Spoiler** tries to prove that the two structures are different.
- **Duplicator** tries to pretend as long as possible, that the structures are the same.

The game is organised in rounds. In each round, Spoiler tries to highlight a difference between the structures by putting an additional pebble in one of the structure. Duplicator tries to match this move in the other structure. The game ends when either the fixed number of rounds is over, or the substructure given by the pebbles differ.

Ehrenfeucht-Fraïssé Games

The game is played on configurations

$$\underbrace{(A, R_1^{\mathcal{A}}, R_2^{\mathcal{A}}, \ldots)}_{\mathcal{A}}, a_1, \ldots, a_k \ \mid \ \underbrace{(B, R_1^{\mathcal{B}}, R_2^{\mathcal{B}}, \ldots)}_{\mathcal{B}}, b_1, \ldots, b_k$$

where $a_1, \ldots, a_k \in A$ and $b_1, \ldots, b_k \in B$.

In each round, Spoiler picks either $a_{k+1} \in A$ or $b_{k+1} \in B$. Then Duplicator picks the other.

Duplicator needs to keep the invariants

- $a_i = a_j$ iff $b_i = b_j$
- $R_i^{\mathcal{A}}(a_{i_1}, \ldots, a_{i_\ell})$ iff $R_i^{\mathcal{B}}(b_{i_1}, \ldots, b_{i_\ell})$

Back to our examples, we see that in the first structure, Duplicator cannot survive a single round. In the second structure, Duplicator can survive the first round, but not the second. And in the third structure, Duplicator can survive two rounds, but not the third.

As mentioned, we're interested in structures $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{A} \equiv_{\mathcal{L}} \mathcal{B}$ for appropriate $\mathcal{L} \subseteq \mathsf{FOL}[\ldots]$. We cannot take $\mathcal{L}$ to be the full of first-order logic, as every finite structure $\mathcal{A} = (\{a_1, \ldots, a_n\}, R_1^{\mathcal{A}}, \ldots)$ can be completely (up to isomorphism) described by a first-order formula.

$$\begin{aligned}
&\exists x_1 \exists x_2 \ldots \exists x_n \\
&\qquad ((\neg)R_1(x_1) \wedge (\neg)R_1(x_2) \wedge \ldots \wedge \ldots \\
&\qquad \wedge \forall y (y = x_1 \vee \ldots \vee y = x_n))
\end{aligned}$$

So we have to restrict the language in some useful way. Also note that it is really the absolute number of quantifiers we have to restrict; just restricting the alternations of $\forall$ and $\exists$, that is restricting the "logical complexity" does not help for finite structures.

We therefore define the "quantifier-rank" as the number of nested (not necessarily alternating) quantifiers a formula contains.

Quantifier-Rank

Definition.

$\mathsf{qr}(\top) = \mathsf{qr}(\bot) = \mathsf{qr}(x = y) = \mathsf{qr}(Rx \ldots z) = 0$

$\mathsf{qr}(\neg\varphi) = \mathsf{qr}(\varphi)$

$\mathsf{qr}(\varphi \wedge \psi) = \mathsf{qr}(\varphi \vee \psi) = \max\{\mathsf{qr}(\varphi), \mathsf{qr}(\psi)\}$

$\mathsf{qr}(\forall x \varphi) = \mathsf{qr}(\exists x \varphi) = \mathsf{qr}(\varphi) + 1$

$\mathsf{FOL}_k[R_1, \ldots, R_n] = \{\varphi \in \mathsf{FOL}[R_1, \ldots, R_n] \mid \mathsf{qr}(\varphi) \leq k\}$

$\mathcal{A}, \vec{a} \equiv_k \mathcal{B}\vec{b}$ iff $\mathcal{A}, \vec{a} \equiv_{\mathsf{FOL}_k[\ldots]} \mathcal{B}, \vec{b}$

The notion of the quantifier-rank turns out to be just the right concept for Ehrenfeucht-Fraïssé Games. An increase of the quantifier rank by one corresponds to another round in the game.

Ehrenfeucht-Fraïssé Theorem

Theorem. For any configuration $\mathcal{A}, \vec{a} \mid \mathcal{B}, \vec{b}$ in an EF-game over finite structures, the following are equivalent.

- Duplicator can survive m more rounds.
- $\mathcal{A}, \vec{a} \equiv_m \mathcal{B}, \vec{b}$

Proof. "⇒"

We actually show the contrapositive: if there is a formula distinguishing $\mathcal{A}, \vec{a}$ and $\mathcal{B}, \vec{B}$ then Spoiler can use this to with the game.

Assume $\mathcal{A}, \vec{a} \models \varphi$, $\mathcal{B}, \vec{b} \not\models \varphi$, $\mathsf{qr}(\varphi) \leq m$.

We argue by induction on m.

If $m = 0$, then Spoiler has won already.

If the two structures differ on an atomic formula, the rules of the game say that Spoiler has won. If they differ on a conjunction or disjunction, then they also differ on one conjunct or disjunct.

So let $m \geq 1$.
W.l.o.g $\varphi \equiv \exists x \psi$.

Concerning the "without loss of generality": if φ is a conjunction, one of the conjuncts would already separate the structures (and this is a simpler formula!); similar for a disjunction. For a negation, just swap the roles of $\mathcal{A}$ and $\mathcal{B}$. If φ is a $\forall x \varphi'$ then $\neg\varphi$ can equivalently written as $\exists x \neg\varphi'$ and this formula obviously has the same quantifier rank.

> Since $\mathcal{A}, \vec{a} \models \exists x \psi$ we can pick a c such that $\mathcal{A}, \vec{a}, c \models \psi$. Spoilers strategy is to place a pebble on c. No matter which d Duplicator picks, it will always be the case that $\mathcal{B}, \vec{b}, d \not\models \psi$. Since $\mathsf{qr}(\psi) + 1 = \mathsf{qr}(\varphi)$ the claim follows by IH.

For the other direction we essentially formalise the rules of the game.

> "$\Leftarrow$"

Winning Condition for $\mathcal{A}, \vec{a}$

Define $\chi_{m,\mathcal{A},\vec{a}}$ s.t. $\mathcal{B}, \vec{b} \models \chi_{m,\mathcal{A},\vec{a}}(\vec{x})$ iff Duplicator has a strategy for m round in $\mathcal{A}, \vec{a} \mid \mathcal{B}, \vec{b}$.

$$\chi_{m+1,\mathcal{A},\vec{a}}(\vec{x}) = (\bigwedge_{a \in A} \exists y \chi_{m,\mathcal{A},\vec{a},a}) \wedge (\forall y \bigvee_{a \in A} \chi_{m,\mathcal{A},\vec{a},a})$$

$$\chi_{0,\mathcal{A},\vec{a}}(\vec{x}) = \bigwedge_{\substack{\mathcal{A}, \vec{a} \models \varphi \\ \varphi \text{ atomic}}} \varphi \quad \wedge \quad \bigwedge_{\substack{\mathcal{A}, \vec{a} \not\models \varphi \\ \varphi \text{ atomic}}} \neg\varphi$$

Note that there are only finitely many $a \in A$ and only finitely many atomic formulae. So we can completely unfold the "big" propositional connectives. Moreover

> $$\mathsf{qr}(\chi_{m,\mathcal{A},\vec{a}}) = m$$

Now, assume $\mathcal{A}, \vec{a} \equiv_m \mathcal{B}, \vec{b}$. Obviously, $\mathcal{A}, \vec{a} \models \chi_{m,\mathcal{A},\vec{a}}$. But then $\mathcal{B}, \vec{b} \models \chi_{m,\mathcal{A},\vec{a}}$. Therefore, Duplicator has a strategy for m rounds.

Example

Let $\mathcal{A} = (\{1, 2, 3\}, <, P_a, P_b)$ be the structure for the word `aab`, i.e. $P_a = \{1, 2\}, P_b = \{3\}$.

- Write down $\chi_{1,\mathcal{A},1}$.
- Does `aaab`, $1 \models \chi_{1,\mathcal{A},1}$?